

Methods in
Molecular Biology 1315

Springer Protocols

Ulrich Lehmann
Jörg Tost *Editors*

Pyrosequencing

Methods and Protocols

Second Edition

 Humana Press

www.Ebook777.com

METHODS IN MOLECULAR BIOLOGY

Series Editor

John M. Walker

School of Life and Medical Sciences

University of Hertfordshire

Hatfield, Hertfordshire, AL10 9AB, UK

For further volumes:

<http://www.springer.com/series/7651>

Pyrosequencing

Methods and Protocols

Second Edition

Edited by

Ulrich Lehmann

Medizinische Hochschule Hannover, Institute of Pathology, Hannover, Germany

Jörg Tost

*Laboratory for Epigenetics and Environment, Centre National de Génotypage, CEA-Institut de Génomique,
Evry, France*

 **Humana Press**

Editors

Ulrich Lehmann
Medizinische Hochschule Hannover
Institute of Pathology
Hannover, Germany

Jörg Tost
Laboratory for Epigenetics and Environment
Centre National de Génotypage
CEA-Institut de Génomique
Evry, France

ISSN 1064-3745

ISSN 1940-6029 (electronic)

Methods in Molecular Biology

ISBN 978-1-4939-2714-2

ISBN 978-1-4939-2715-9 (eBook)

DOI 10.1007/978-1-4939-2715-9

Library of Congress Control Number: 2015939912

Springer New York Heidelberg Dordrecht London

© Springer Science+Business Media New York 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Humana Press is a brand of Springer

Springer Science+Business Media LLC New York is part of Springer Science+Business Media (www.springer.com)

Preface

The Pyrosequencing technology has been around for nearly 20 years and has become a widely used technology for a broad variety of applications to analyze genetic and epigenetic variation as well as to identify bacteria, viruses, and microorganisms. Pyrosequencing is a versatile, flexible, accurate, and extremely quantitative sequencing-by-synthesis technology, which is relatively easy to handle and has a short time to results. The sequence-based read-out, and the rapid and standardized work-flow of Pyrosequencing as well as the improved limit of detection and analytical sensitivity compared to traditional sequencing technologies, make this technology a method of choice for diagnostic applications. Furthermore, it has established itself as the “gold standard” platform for quantitative locus-specific DNA methylation analysis.

Eight years have passed since the first edition of this volume edited by Sharon Marsh in 2007 and genomics and medical diagnostics have undergone a major revolution in the time from the last edition. While Pyrosequencing has been used as the underlying principle for the first commercially available next-generation sequencing system (454 Life Sciences, now part of Roche Molecular Diagnostics), the technology has not been able to evolve as rapid as other sequencing platforms. However, its use in locus-specific analyses is still expanding and novel applications are continuously devised. Despite the rapid advancements in the field of comprehensive high-throughput genome-wide “omics” technologies, declining costs, and the development of more user-friendly systems, we are convinced that fast, cost-effective, focused, robust, and widely distributed technologies with a moderate throughput, like Pyrosequencing, will have their place for the foreseeable future, not only for the independent validation of findings from genome-wide screening approaches but also for the exact quantitative analysis of single (or only a few) loci and within the molecular diagnostic laboratory.

The second edition of the “Pyrosequencing Protocols” contains 27 exciting chapters written by experts in the field that use this platform for their different analyses. With the exception of the chapter describing the history of Pyrosequencing and its development by Pal Nyren, one of the inventors of the technology who updated his historical overview, all other chapters have been newly written for this edition and we hope to cover a broad range of application. Chapters 1–4 provide a general overview on the biochemistry, the available instruments, tools for the interpretation of the sequencing output (Pyrograms), as well as general considerations on the strength and limitations of the technology that are of interest to all applications of the technology. Chapters 5–12 focus on the quantitative analysis of genetic variation, including detection of mutations of clinical relevance in different biological specimens (Chapters 5–8), copy number variation (Chapter 9), and large-scale chromosomal alterations (Chapter 10). Chapter 11 describes the specific genotyping of HLA alleles using Pyrosequencing, while Chapter 12 investigates the ratio of expressed alleles at the RNA level. Chapters 13–22 focus on the analysis of DNA methylation including gene-specific (Chapters 13 and 14) and global DNA methylation assays (Chapters 15 and 16). Specialized applications for DNA methylation analysis described in this volume include

single molecule analysis (Chapter 17), loss of imprinting (Chapter 18), single blastocyst analysis (Chapter 19), allele-specific DNA methylation patterns (Chapters 20 and 21), as well as the analysis of DNA methylation patterns associated with specific histone modifications (Chapter 22). Tools and protocols for the detection of viruses and bacteria are described in Chapters 23–25, while the exciting possibilities of genetic and epigenetic analyses for forensics using Pyrosequencing are described in Chapters 26 and 27.

This volume of the “Methods in Molecular Biology” series is addressed to advanced doctoral students and postdoctoral investigators and research scientists that are implicated in the different aspects of genetics and cellular and molecular biology as well as to clinicians involved in diagnostics or choice of treatment of diseases that have an analytical component. The presentation in this volume is equally suited for laboratories that already have a great deal of expertise in Pyrosequencing, but might want to extend the use of this technology to other applications, and for genetics/genomics/biology groups that want to initiate research projects making use of the advantages of the Pyrosequencing technology. Notes and tips from the experts for the different methods will enable a rapid implementation of the different protocols in the laboratory and avoid time-consuming and cost-intensive mistakes.

We are indebted to all the authors for their hard work and outstanding contributions to this second edition of “Pyrosequencing Protocols.” It was a pleasure to work with them on this project. The series editor John Walker provided essential initial guidance, and the staff from Springer was helpful in finalizing the publication process. We would also like to thank Britta Hasemeier for her great help in harmonizing all artwork and reference lists produced by more than 20 different groups.

We sincerely hope that the protocols described in detail in this volume will enable to develop new applications of the Pyrosequencing technology increasing further the great versatility of the highly accurate and quantitative technology and will help to enhance our understanding of the molecular processes underlying biological processes and disease for the benefit of patients.

Hannover, Germany
Evry, France

Ulrich Lehmann
Jörg Tost

Contents

<i>Preface</i>	<i>v</i>
<i>Contributors</i>	<i>xi</i>

PART I INTRODUCTION

1 The History of Pyrosequencing®	3
<i>Pål Nyrén</i>	
2 PyroMark® Instruments, Chemistry, and Software for Pyrosequencing® Analysis.	17
<i>Martin Kreutz, Gerald Schock, Julia Kaiser, Norbert Hochstein, and Ralf Peist</i>	
3 Software-Based Pyrogram® Evaluation.	29
<i>Guoli Chen, Matthew T. Olson, and James R. Eshleman</i>	
4 Quantitative Validation and Quality Control of Pyrosequencing® Assays	39
<i>Ulrich Lehmann</i>	

PART II ANALYSIS OF GENETIC VARIATION

5 Extended <i>KRAS</i> and <i>NRAS</i> Mutation Profiling by Pyrosequencing®	49
<i>Andreas Jung</i>	
6 Universal BRAF State Detection by the Pyrosequencing®-Based U-BRAF ^{V600} Assay.	63
<i>Alexander Skorokhod</i>	
7 Pyrosequencing®-Based Identification of Low-Frequency Mutations Enriched Through Enhanced- <i>ice</i> -COLD-PCR	83
<i>Alexandre How-Kit and Jörg Tost</i>	
8 Analysis of Mutational Hotspots in Routinely Processed Bone Marrow Trephines by Pyrosequencing®	103
<i>Stephan Bartels and Ulrich Lehmann</i>	
9 Analysis of Copy Number Variation by Pyrosequencing® Using Paralogous Sequences	115
<i>Marianne Kristiansen Kringen</i>	
10 Prenatal Diagnosis of Chromosomal Aneuploidies by Quantitative Pyrosequencing®	123
<i>Hui Ye, Haiping Wu, Yunlong Liu, Bingjie Zou, Tomoharu Kajiyama, Hideki Kambara, and Guohua Zhou</i>	
11 HLA-B and HLA-C Supratyping by Pyrosequencing®	133
<i>Irene Vanni, Elisabetta Ugolotti, Patrizia Larghero, and Roberto Biassoni</i>	

- 12 Allele Quantification Pyrosequencing® at Designated SNP Sites to Detect Allelic Expression Imbalance and Loss-of-Heterozygosity 153
Chau-To Kwok and Megan P. Hitchins

PART III ANALYSIS OF DNA METHYLATION PATTERNS

- 13 Quantitative DNA Methylation Analysis by Pyrosequencing® 175
Jessica Roessler and Ulrich Lehmann
- 14 Quantitative Methylation Analysis of the *PCDHB* Gene Cluster 189
Barbara Banelli and Massimo Romani
- 15 Assessment of Changes in Global DNA Methylation Levels by Pyrosequencing® of Repetitive Elements 201
Ali M. Tabish, Andrea A. Baccarelli, Lode Godderis, Timothy M. Barrow, Peter Hoet, and Hyang-Min Byun
- 16 Global Analysis of DNA 5-Methylcytosine Using the Luminometric Methylation Assay, LUMA 209
Karin Luttröpp, Louise K. Sjöholm, and Tomas J. Ekström
- 17 Limiting Dilution Bisulfite Pyrosequencing®: A Method for Methylation Analysis of Individual DNA Molecules in a Single or a Few Cells 221
Nady El Hajj, Julianne Kubtz, and Thomas Haaf
- 18 Detection of Loss of Imprinting by Pyrosequencing® 241
Silvia Tabano, Eleonora Bonaparte, and Monica Miozzo
- 19 Analysis of DNA Methylation Patterns in Single Blastocysts by Pyrosequencing® 259
John Huntriss, Kathryn Woodfine, Joanna E. Huddleston, Adele Murrell, and Helen M. Picton
- 20 Allele-Specific DNA Methylation Detection by Pyrosequencing® 271
Lasse Sommer Kristensen, Jens Vilstrup Johansen, and Kirsten Grønbaek
- 21 SNP-based Quantification of Allele-specific DNA Methylation Patterns by Pyrosequencing® 291
Florence Busato and Jörg Tost
- 22 DNA Methylation Analysis of ChIP Products at Single Nucleotide Resolution by Pyrosequencing® 315
Céline Moison, Fanny Assemat, Antoine Daunay, Paola B. Arimondo, and Jörg Tost

PART IV BACTERIAL TYPING AND IDENTIFICATION

- 23 Multiplex Pyrosequencing®: Simultaneous Genotyping Based on SNPs from Distant Genomic Regions 337
Piotr Wojciech Dabrowski, Kati Bourquain, and Andreas Nitsche
- 24 Detection of Drug-Resistant *Mycobacterium tuberculosis* 349
Anna Engström and Pontus Jureén
- 25 Application of Pyrosequencing® in Food Biodefense 363
Kingsley Kwaku Amoako

PART V FORENSICS

26	Forensic Analysis of Mitochondrial and Autosomal Markers Using Pyrosequencing®	379
	<i>Magdalena M. Buś, Hanna Edlund, and Marie Allen</i>	
27	Tissue-Specific DNA Methylation Patterns in Forensic Samples Detected by Pyrosequencing®	397
	<i>Joana Antunes, Kuppareddi Balamurugan, George Duncan, and Bruce McCord</i>	
	<i>Index</i>	411

Contributors

- MARIE ALLEN • *Department of Immunology, Genetics, and Pathology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden*
- KINGSLEY KWAKU AMOAKO • *National Centers for Animal Disease, Lethbridge Laboratory, Canadian Food Inspection Agency, Lethbridge, AB, Canada*
- JOANA ANTUNES • *Department of Chemistry and Biochemistry, Florida International University, Miami, FL, USA*
- PAOLA B. ARIMONDO • *Unité de Service et de Recherche CNRS-Pierre Fabre n°3388, Epigenetic Targeting of Cancer (ETaC), Toulouse, France*
- FANNY ASSEMAT • *Unité de Service et de Recherche CNRS-Pierre Fabre n°3388, Epigenetic Targeting of Cancer (ETaC), Toulouse, France*
- ANDREA A. BACCARELLI • *Laboratory of Environmental Epigenetics, Exposure Epidemiology and Risk Program, Harvard School of Public Health, Boston, MA, USA*
- KUPPAREDDI BALAMURUGAN • *School of Criminal Justice, The University of Southern Mississippi, Hattiesburg, MS, USA*
- BARBARA BANELLI • *Laboratory of Tumor Epigenetics, IRCCS AOU San Martino–IST, Genoa, Italy*
- TIMOTHY M. BARROW • *Institute for Prevention and Tumor Epidemiology, Freiburg Medical Center, University of Freiburg, Freiburg, Germany; German Consortium for Translational Cancer Research (DKTK), Heidelberg, Germany*
- STEPHAN BARTELS • *Institute of Pathology, Medizinische Hochschule Hannover, Hannover, Germany*
- ROBERTO BIASSONI • *Molecular Medicine, Translational Medicine Department, Istituto Giannina Gaslini, Genoa, Italy*
- ELEONORA BONAPARTE • *Division of Pathology, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy*
- KATI BOURQUAIN • *biotechrabbit GmbH, Berlin, Germany*
- MAGDALENA M. BUŚ • *Department of Immunology, Genetics, and Pathology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden*
- FLORENCE BUSATO • *Laboratory for Epigenetics and Environment, Centre National de Génotypage, CEA-Institut de Génomique, Evry, France*
- HYANG-MIN BYUN • *Laboratory of Environmental Epigenetics, Exposure Epidemiology and Risk Program, Harvard School of Public Health, Boston, MA, USA*
- GUOLI CHEN • *Department of Pathology, Penn State Milton S. Hershey Medical Center, Hershey, PA, USA*
- PIOTR WOJCIECH DABROWSKI • *ZBS 1: Highly Pathogenic Viruses/ZV4: Information Technology, Robert Koch Institute, Berlin, Germany*
- ANTOINE DAUNAY • *Laboratory for Functional Genomics, Fondation Jean Dausset–CEPH, Paris, France*
- GEORGE DUNCAN • *Department of Chemistry and Biochemistry, Florida International University, Miami, FL, USA*

- HANNA EDLUND • *Department of Immunology, Genetics, and Pathology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden*
- TOMAS J. EKSTRÖM • *Department of Clinical Neuroscience, Karolinska Institutet, Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden*
- ANNA ENGSTRÖM • *Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden; Molecular Mycobacteriology, Research Centre Borstel, Borstel, Germany*
- JAMES R. ESHLEMAN • *Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD, USA*
- LODE GODDERIS • *Department of Public Health and Primary Care, Katholieke Universiteit Leuven, Leuven, Belgium; IDEWE, External Service for Prevention and Protection at work, Heverlee, Belgium*
- KIRSTEN GRØNBÆK • *Department of Hematology, Copenhagen, Denmark*
- THOMAS HAAF • *Institute of Human Genetics, Julius-Maximilians University Würzburg, Würzburg, Germany*
- NADY EL HAJJ • *Institute of Human Genetics, Julius-Maximilians University Würzburg, Würzburg, Germany*
- MEGAN P. HITCHINS • *Stanford University School of Medicine, Stanford Cancer Institute, Stanford, CA, USA*
- NORBERT HOCHSTEIN • *QIAGEN GmbH, Hilden, Germany*
- PETER HOET • *Department of Public Health and Primary Care, Katholieke Universiteit Leuven, Leuven, Belgium*
- ALEXANDRE HOW-KIT • *Laboratory for Functional Genomics, Fondation Jean Dausset-CEPH, Paris, France*
- JOANNA E. HUDDLESTON • *Department of Oncology, Cancer Research UK Cambridge Research Institute, University of Cambridge, Cambridge, UK*
- JOHN HUNTRISS • *Division of Reproduction and Early Development, Leeds Institute of Cardiovascular and Molecular Medicine, University of Leeds, Leeds, UK*
- JENS VILSTRUP JOHANSEN • *Biotech Research and Innovation Centre, University of Copenhagen, Copenhagen, Denmark*
- ANDREAS JUNG • *Pathologisches Institut, Ludwig-Maximilians-Universität, Munich, Germany*
- PONTUS JURÉEN • *Unit for Highly Pathogenic Microorganisms, Department of Microbiology, Public Health Agency of Sweden, Solna, Sweden*
- JULIA KAISER • *QIAGEN GmbH, Hilden, Germany*
- TOMOHARU KAJIYAMA • *Central Research Laboratory, Hitachi, Ltd., Kokubunji-shi, Tokyo, Japan*
- HIDEKI KAMBARA • *Central Research Laboratory, Hitachi, Ltd., Kokubunji-shi, Tokyo, Japan*
- MARTIN KREUTZ • *QIAGEN GmbH, Hilden, Germany*
- MARIANNE KRISTIANSEN KRINGEN • *Department of Pharmacology, Oslo University Hospital, Oslo, Norway*
- LASSE SOMMER KRISTENSEN • *Department of Hematology, Copenhagen, Denmark*
- JULIANE KUHTZ • *Institute of Human Genetics, Julius-Maximilians University Würzburg, Würzburg, Germany*
- CHAU-TO KWOK • *Adult Cancer Program, Lowy Cancer Research Centre, University of New South Wales, Sydney, NSW, Australia*

- PATRIZIA LARGHERO • *Molecular Medicine, Translational Medicine Department, Istituto Giannina Gaslini, Genoa, Italy*
- ULRICH LEHMANN • *Institute of Pathology, Medizinische Hochschule Hannover, Hannover, Germany*
- YUNLONG LIU • *Department of Pharmacology, Jinling Hospital, Medical School, Nanjing University, Nanjing, PR China*
- KARIN LUTTROPP • *Department of Molecular Medicine and Surgery, Karolinska Institutet, Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden*
- BRUCE MCCORD • *Department of Chemistry and Biochemistry, Florida International University, Miami, FL, USA*
- MONICA MIOZZO • *Department of Pathophysiology and Transplantation, Università degli Studi di Milano, Milan, Italy; Division of Pathology, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy*
- CÉLINE MOISON • *Unité de Service et de Recherche CNRS-Pierre Fabre n°3388, Epigenetic Targeting of Cancer (ETaC), Toulouse, France; Université Pierre et Marie Curie, Paris, France*
- ADELE MURRELL • *Biology and Biochemistry Department, Centre for Regenerative Medicine, University of Bath, Bath, UK*
- ANDREAS NITSCHKE • *ZBS 1: Highly Pathogenic Viruses/ZV4: Information Technology, Robert Koch Institute, Berlin, Germany*
- PÅL NYRÉN • *Department of Biotechnology, Royal Institute of Technology, Stockholm, Sweden*
- MATTHEW T. OLSON • *Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD, USA*
- RALF PEIST • *QIAGEN GmbH, Hilden, Germany*
- HELEN M. PICTON • *Division of Reproduction and Early Development, Leeds Institute of Cardiovascular and Molecular Medicine, University of Leeds, Leeds, UK*
- JESSICA ROESSLER • *Institute of Pathology, Medizinische Hochschule Hannover, Hannover, Germany*
- MASSIMO ROMANI • *Laboratory of Tumor Epigenetics, IRCCS AOU San Martino - IST, Genoa, Italy*
- GERALD SCHOCK • *QIAGEN GmbH, Hilden, Germany*
- LOUISE K. SJÖHOLM • *Department of Clinical Neuroscience, Karolinska Institutet, Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden*
- ALEXANDER SKOROKHOD • *Department of Dermatology, Heidelberg University Hospital, Heidelberg, Germany*
- SILVIA TABANO • *Department of Pathophysiology and Transplantation, Università degli Studi di Milano, Milan, Italy; Division of Pathology, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy*
- ALI M. TABISH • *Department of Public Health and Primary Care, Katholieke Universiteit Leuven, Leuven, Belgium*
- JÖRG TOST • *Laboratory for Epigenetics and Environment, Centre National de Génotypage, CEA-Institut de Génomique, Evry, France*
- ELISABETTA UGOLOTTI • *Molecular Medicine, Translational Medicine Department, Istituto Giannina Gaslini, Genoa, Italy*
- IRENE VANNI • *Molecular Medicine, Translational Medicine Department, Istituto Giannina Gaslini, Genoa, Italy*

KATHRYN WOODFINE • *Department of Oncology, Cancer Research UK Cambridge Research Institute, University of Cambridge, Cambridge, UK*

HAIPING WU • *Huadong Research Institute for Medicine and Biotechnics, Nanjing, PR China*

HUI YE • *Department of Pharmacology, Jinling Hospital, Medical School, Nanjing University, Nanjing, PR China*

GUOHUA ZHOU • *Department of Pharmacology, Jinling Hospital, Medical School, Nanjing University, Nanjing, PR China*

BINGJIE ZOU • *Department of Pharmacology, Jinling Hospital, Medical School, Nanjing University, Nanjing, PR China*

Part I

Introduction

Free ebooks ==> www.Ebook777.com

Chapter 1

The History of Pyrosequencing®

Pål Nyrén

Abstract

One late afternoon in the beginning of January 1986, bicycling from the lab over the hill to the small village of Fulbourn, the idea for an alternative DNA sequencing technique came to my mind. The basic concept was to follow the activity of DNA polymerase during nucleotide incorporation into a DNA strand by analyzing the pyrophosphate released during the process. Today, the technique is used in multidisciplinary fields in academic, clinical, and industrial settings all over the world. This technique can be used for both single-base sequencing and whole-genome sequencing, depending on the format used.

In this chapter, I give my personal account of the development of Pyrosequencing®—beginning on a winter day in 1986, when I first envisioned the method—until today, nearly 30 years later.

Key words Pyrosequencing®, Sequencing, Bioluminescence, Pyrophosphate

1 Background

Pyrosequencing® is a DNA sequencing technique that utilizes enzyme-coupled reactions and bioluminescence to monitor the pyrophosphate release accompanying nucleotide incorporation, in real time. The Pyrosequencing method [1, 2] was the first available commercial alternative to the well-known Sanger method for de novo DNA sequencing. Several thousands of scientific papers comprise the literature describing the development and applications of Pyrosequencing (<http://www.ncbi.nlm.nih.gov/pubmed/?term=pyrosequencing>). This method can be used for single nucleotide polymorphism analysis and tag sequencing (up to 100 bases), as well as for whole-genome sequencing [3].

It was during my postdoctoral period in Cambridge when it occurred to me that the detection of pyrophosphate could be used to sequence DNA. While working with Sir John Walker (later awarded the 1997 Nobel Prize for Chemistry) at the Medical Research Council Laboratory of Molecular Biology (LMB), one of my projects was to isolate and sequence the gene for the bovine mitochondrial phosphate-carrier protein [4]. Sanger sequencing

was performed manually, using radioactive-labeled nucleotides and handmade thin gels for detection and separation. Sequencing was an arduous business at the time, requiring weeks of practice to learn the procedure. Several steps were involved in the method, leaving time for the mind to wander in between. As a newcomer, I was not always successful with the handling of very thin acrylamide gels. I remember thinking that it would be great if the method could be simplified or if sequencing could be performed in some other way. During my time as Ph.D. student at Stockholm University, I worked extensively with modification and simplification of a variety of methods, as well as with the development of new procedures, so it was natural for me to look for ways to improve methods that I found cumbersome and labor intensive.

One late afternoon in the beginning of January 1986, bicycling from the lab over the hill to the small village of Fulbourn, the idea for an alternative DNA sequencing technique came to my mind. It was late, dark, and rainy as I hurried home to tell my wife Maija about the new idea. She later told me that when I explained the new idea to her, she thought that I looked like Gyro Gearloose's little helper—the bright-headed assistant with a light bulb as a head. I had difficulty in sleeping that night and was eager to go home to Sweden to test my new idea. What I could not expect that day was that 10 years would pass before the method was fully developed.

The basic concept was to follow the activity of DNA polymerase during nucleotide incorporation into a DNA strand by analyzing the pyrophosphate released during the process. Why pyrophosphate detection? As a Ph.D. student, I worked in the fields of bioenergetics and photosynthesis. My Ph.D. project was to isolate and study an enzyme endogenous to a photosynthetic bacterium. The enzyme, proton-translocating inorganic pyrophosphatase, is involved in photophosphorylation and catalyzes light-driven pyrophosphate synthesis. In order to follow the light-driven process, I developed a very sensitive luminometric method [5] capable of following pyrophosphate synthesis in real time, both after continuous illumination [6] and after short light flashes [7]. This was a break-through in bioenergetics research. The new idea was to follow the DNA polymerase activity using the pyrophosphate method, thereby detecting whether or not a base is incorporated during DNA synthesis. Because Watson-Crick base-pairing rules that nucleotide G will always pair with C and T with A, DNA polymerase activity will decipher the base composition of the template when known nucleotides are utilized.

As stated previously, it required nearly 10 years to get the sequencing method working, involving much struggle to get financing and support for my idea, and to solve several problems that turned up as the project evolved. In the following sections, I will describe the most important parts of the developing process.

2 Envisioning Pyrosequencing

As I previously mentioned, I wanted to go back to Sweden as quickly as possible to test my new idea in the lab. However, we could not leave Cambridge before my postdoctoral period was completed. Of course I also wanted to accomplish something significant with my work at LMB before leaving. To expedite the process, I wrote a letter to my friend Åke Strid, asking him to advise a new student to purchase the necessary chemicals and set up a pilot study of the DNA polymerase assay. When I called after a month to check how things were going, he informed me that he could not get the suggested assay to work. I had to wait until the end of November 1986 before I could investigate what went wrong with my idea.

Back in Stockholm, I applied for research funding for two projects: my old bioenergetics project and the new sequencing concept. Unfortunately, I obtained financing only for the old project. The committee did not share my enthusiasm for the new idea, reasoning that ATP is a substrate for DNA synthesis and therefore would interfere with the luciferase assay. However, this point is wrong; ATP is not a substrate for DNA polymerase. The head of the department at that time had already advised me (without reading my project plans) not to apply for two separate projects because of the tough competition. Later, when I had already moved to the Royal Institute of Technology (KTH), and the project was shown to be successful, he alleged that it was a pity that I did not have the idea during my time at the Stockholm University.

Although I did not get research money for the DNA sequencing project, I decided to use a small portion of money from the other project to buy a few necessary chemicals and enzymes. I then spent evenings and weekends developing the DNA polymerase assay [8]. I eventually found that the method worked. I do not know why it had not worked for the student. I have learned that it is always better to do things yourself, if possible, rather than try to persuade another person, who might already be distracted by other projects. Even after the paper was published, I continued to apply for research funding without success. The project was shelved and I had to wait several years before I could continue to develop the method for use in DNA sequencing.

Much later, I learned that Bob Melamede [9], whom I met in Stockholm in 1997, had described the general principles of DNA sequencing-by-synthesis in a previously obtained patent. The method was based on detection of the decrease in nucleotide absorbance upon nucleotide incorporation. The method's sensitivity was not sufficient for normal DNA concentrations. The pyrophosphate approach had an important advantage in terms of sensitivity; DNA quantities obtained by standard PCR procedures

are adequate for the analysis with this firefly luciferase-based method. The sensitivity with absorbance difference measurements is probably 100–1,000 times lower. When I later met Bob, he was very happy to hear that his sequencing-by-synthesis concept worked and that I had circumvented the problem of DNA polymerase activity monitoring.

3 Solid-Phase Pyrosequencing and the α -S-Nucleotide

In 1990, I moved to KTH in Stockholm to join Professor Jan Rydstöm's group. Together with Åke Strid, studies of the bovine membrane bound transhydrogenase were initiated, and we were also able to continue our old projects in Rydström's lab. As a newcomer at KTH, I spent some time reading old theses and articles published by the Biochemistry department. One especially attracted my interest: it described a method for solid-phase DNA sequencing [10]. Paramagnetic beads were utilized for DNA template preparation before sequencing. I theorized that if I could combine the magnetic bead technique with my DNA analysis system, a DNA sequencing procedure would be possible. I consulted Professor Mathias Uhlén, who was the principal investigator for the magnetic bead project, and shortly thereafter I collaborated with Bertil Pettersson, a student from Mathias Uhlén's group. Bertil taught me everything that I needed to know about working with magnetic beads. He synthesized oligonucleotides and prepared DNA by PCR. I worked with the magnetic beads and the pyrophosphate detection system. During the first 4 years I worked 1–2 days a week with the new project.

A problem I faced during the initial stages of the project was high background during the luminescence measurement. The enzymes and the nucleotides were the source of most of the background signals. One approach I used to lower the background was preincubation with apyrase immobilized on magnetic beads; the method was published several years later [11]. I also used to pretreat the nucleotides with pyrophosphatase to decrease the pyrophosphate content. Today, this is a standard procedure. The most severe problem I observed was that dATP functioned as a weak substrate for luciferase giving light corresponding to 2–3 % of an equal amount of ATP. I mediated this difficulty by starting all reactions by adding DNA polymerase and not by adding nucleotides. This procedure made it unnecessary to subtract a high background from a low signal. Because most polymerases were delivered in phosphate buffers, these enzymes also had to be pretreated with pyrophosphatase.

Although solutions to several of the mentioned problems had now been found, I still had difficulty getting strong, clear signals. Bertil then came up with the brilliant proposal of setting up the

analysis system quite differently. We used run-off signals to increase signal intensity, and dideoxynucleotides were used to facilitate reading the first base position. This new method was the first proof-of-principle for the DNA sequencing concept, and the success was encouraging for future development work [12]. Later we published an alternative method for single-base change detection [13]. The concept relies on measuring the differences in primer extension efficiency by a DNA polymerase of a matched over a mismatched 3'-terminal utilizing α -thiotriphosphate analogs. One important feature of the DNA sequencing concept is that DNA polymerases lacking 3'-5' activity are required, which limited us to the use of Sequenase and a modified Klenow enzyme. When we used dideoxynucleotides, we could only use Sequenase due to their hindrance of Klenow's activity.

To be able to sequence multiple bases we had to work on decreasing the dATP background. I made a literature search for alternative nucleotides and decided to test an α -S-modified nucleotide. Nucleoside thiophosphates comprised a new class of modified nucleotides in which one nonbridging oxygen atom in the α -phosphate of the nucleoside 5'-triphosphate is replaced by a sulfur group. I found that the modified nucleotide was a good substrate for the polymerase and in addition, a poor substrate for luciferase. This finding was made during 1995 and was published 1996 [14]. We were able to sequence 15 bases with our new approach—a new world record for the sequencing-by-synthesis principle.

From 1991 to 1994, I was engaged in several other projects [15–17], and the development speed for the DNA sequencing method was low. It was essentially a one-man project with little or no funding, and I spent about 25 % of my time with research and the rest with teaching obligations. From 1994 to 1998, my one-man group increased to several people, at most eight, working with different aspects of the DNA method. No one was particularly enthusiastic about the new idea during the earlier period—when I presented my idea at a conference in Stockholm, very few people showed interest in my poster. However, one well-known scientist (an adviser for the government) approached my poster, looked it over, and then turned to me, saying, “I don't think this will ever work.” That comment made me a bit anxious, as she was an expert in nucleotide metabolism. Regardless, I continued to believe in my idea.

Up to this point, we had used magnetic beads and manual sequencing, but our aim was to automate the procedure. Together with Professor Johan Roeraade, Ph.D., Sean Waters, and my students Mostafa Ronaghi, Tommy Nordström, and Atefeh Shakri, I started to explore the possibility of immobilizing enzymes and DNA on silica with the goal of constructing an automated sequencing procedure utilizing a capillary flow system. At that time, we also

started a collaboration with Björn Ekström at Pharmacia Biotech. We were able to immobilize both luciferase and ATP sulfurylase and successfully analyze ATP and pyrophosphate continuously in the capillary system (unpublished results). We also made preliminary studies of nucleotide incorporations on immobilized DNA.

4 Apyrase

While attempting to solve problems such as low-signal intensity and nucleotide contamination associated with the capillary flow system, another idea occurred to me. Instead of including a washing step between the nucleotide additions, it might be possible to utilize a nucleotide-degrading enzyme. My first thought was to use my earlier published [11] concept of apyrase immobilized on magnetic beads. The immobilized apyrase could be separated from the assay after each degradation step by simply using a magnet. Alternatively, if the kinetics were properly adjusted, it might be possible to omit the separation step altogether. However, it remained to be seen whether apyrase could degrade nucleotides other than ATP and if all four deoxynucleotides were equally well degraded. At that time, no one was especially impressed or excited with this proposal. I asked one of my students to test the concept, but he later told me that it did not work. I could not let the idea go, so during the summer 1996 vacancies I went back to the lab bench and started to set up some really exciting experiments.

After about 6 weeks in the laboratory, I had preliminary data for a new DNA sequencing procedure ready. The main obstacle that I had to handle was false signals, which I theorized were a result of nucleoside diphosphate kinase activity. I identified the ATP sulfurylase preparation as the main contamination source for this activity. As predicted, the false signals decreased with lower concentrations of ATP sulfurylase. Two months later, I had evidence supporting my theory [18]. I started using a new source of ATP sulfurylase purified from dry yeast (bought from a nearby food store) by my student Nader Nourizad. Another student, Samer Karamohamed, was later able to produce a recombinant form of the enzyme in *Escherichia coli* [19]. Both the purified enzyme and the recombinant ATP sulfurylase improved the sequencing results dramatically. Together with Ronaghi, I started to optimize all parameters of the method, and in early 1998 the protocol was ready [1, 2]. Only minor changes have since been made to the protocol.

One problem we encountered during the optimization process was that some templates produced significantly better results than others. We hypothesized that some DNA templates formed secondary structures because of the relatively low temperature (22 °C) that was used. We tried to solve this issue by adding different

substances such as glycerol, proline, and DMSO, but none of these cheaper substances helped with this problem. Ronaghi then suggested that we should test the effects of adding a single-stranded DNA-binding protein (SSB). Most organisms utilize this protein to decrease secondary-structure formation, thereby improving DNA synthesis. I had earlier found that SSB could be used to improve DNA priming with two or three unligated hexamers, so the substance was still available in our lab. SSB was a hit; it dramatically improved the sequence quality for difficult templates and for less difficult templates to some degree as well [20, 21].

5 Automation

With the aim of commercializing the new DNA sequencing technique, the company Pyrosequencing AB was founded in 1997 by Pål Nyrén, Mathias Uhlén, Mostafa Ronaghi, Bertil Pettersson, and Björn Ekström. The first commercially available automated system was sold in 1999. The first noncommercially available automated system was developed by myself, Tommy Nordström, and Mostafa Ronaghi [22]. It included an LKB-1251 luminometer, two dispenser controllers, a power unit, a computer, a recorder, and four separate dispensers. At 1–5 μL , the dispensed volume was rather large, as was the sample volume at 0.4 mL. The system was also restricted to sequencing one sample at a time. In contrast, the commercial systems developed later sequence 96 samples in parallel, and both dispensation volumes (0.05–0.2 μL) and sample volumes (10–50 μL) are much smaller, substantially decreasing the cost and time for a DNA sequencing project.

6 Improvements

In addition to the development process, several aspects of the Pyrosequencing method have been improved during the following years. I will briefly mention a few of the most important alterations. We had observed major decreases in apyrase activity following the substitution of an S-modified variant for the standard dATP during longer sequencing projects. The new nucleotide used was a mixture of two isomers—one of which, the R-isomer, is not a substrate for the DNA polymerase. We therefore decided to test a nucleotide solution consisting of only the active S-isomer. By utilizing a pure isomer, we could decrease the nucleotide concentration by one-half and thereby dramatically decrease the apyrase inhibition. We obtained much longer reads using the new approach [23]; up to 153 bases of sequence information could be analyzed on one of the studied templates.

It was obvious from the above experiment that the modified A nucleotide had a negative effect on apyrase activity. Although the problem was decreased by the use of the pure isomer, we could still observe apyrase inhibition, especially during long sequencing projects. After a deep literature search, I came to the conclusion that only a few nucleotide analogs were potentially appropriate for the Pyrosequencing method. We tested 7-deaza-2'-deoxyadenosine-5'-triphosphate and observed positive effects on the sequencing data. The nucleotide was incorporated by the polymerase with good efficiency, and in addition it had no negative effect on apyrase activity even after many nucleotide additions [24]. We surmised that this nucleotide might be a good choice for use in future Pyrosequencing protocols.

Another improvement of the sequencing method, especially for templates with homopolymeric T-regions, has been the introduction of Sequenase. For example, a template containing an 8-mer poly(T) track could be easily read when Sequenase was used, but not when Klenow polymerase was used [25]. Sequenase could also be used in combination with 7-deaza-2'-deoxyadenosine-5'-triphosphate [24].

We had earlier shown that SSB improved the sequence quality for some templates but the effect of temperature had not been investigated due to the thermo-instability of firefly luciferase. By introducing glycine betaine, we were able to increase the temperature for the sequencing procedure by approximately 10 °C, up to 37 °C [26, 27]. At this higher temperature, the activity of the different enzymes was doubled and the sequence quality for some templates improved.

Template preparation for Pyrosequencing has been substantially improved by the introduction of a Sepharose® bead/vacuum system that allows preparation of 96 samples within a few minutes. Another strategy, introduced by Tommy Nordström, allows double-stranded DNA to be used as a template for DNA sequencing after a simple preparation procedure [28–30]. Using a combination of apyrase, pyrophosphatase, and blocking oligonucleotides, the template can be prepared by a one-step procedure within a few minutes [30].

Preprogrammed dispensation order was introduced to improve read length [31]. This strategy has applications in resequencing projects, such as mutation detection in cancer research and verification of clones [31, 32].

A major improvement of the Pyrosequencing method was the introduction of the multiprimer DNA sequencing concept by my student Baback Gharizadeh [33–35]. In this method, two or more sequencing primers, combined in a pool, are added to a DNA sample of interest. The oligonucleotides that hybridize to the DNA sample will function as primers during the subsequent DNA sequencing procedure. This new strategy is suitable for sequencing and

typing of samples harboring different genotypes (coinfections with multiple genotypes) and samples yielding nonspecific amplifications. This therefore eliminates the need for nested PCR, stringent PCR conditions, and cloning. The new approach has also proved to be useful for amplicons containing low yields or subdominant types.

7 Applications

DNA sequencing has become an invaluable tool in disparate fields such as medicine, agriculture, and forensic studies. Pyrosequencing can be used for both single-base sequencing and whole-genome sequencing, depending on the format used. The developed picotiterplate format allows whole-genome sequencing with very high throughput: a 100-fold increase over current Sanger sequencing technology [3, 36]. The conventionally used microtiter plate format is used to sequence 96 samples in parallel for up to or more than 100 bases [23, 34]. Novel formats include a pyrosequencer analyzing 24 samples in parallel (*see* Chapter 2). Earlier used primarily for SNP and mutation analyses [28, 37] (*see* Chapters 5–8 and 10), the method now has numerous different applications: analysis of allele frequency in pooled samples [38], DNA methylation analyses [39] (*see* Chapters 13–22), molecular haplotyping [40], sequencing of heteroplasmic DNA [41], forensic analyses (*see* Chapters 26 and 27), bacterial typing [34] (*see* Chapters 23–25), fungal typing [45], and viral typing [46]. Furthermore, Pyrosequencing facilitates clinical research in areas such as Alzheimer's disease, autoimmune disorders, bowel disease, cardiovascular disease, coronary heart disease, dermatology, diabetes, gynecology, hematology, hearing loss, hematopoietic chimerism, immunology, mitochondrial disorders, nephrology, neurology, obesity, oncology, orthopedics, psychiatric genetics, and trauma (for references see <http://www.ncbi.nlm.nih.gov/pubmed/?term=pyrosequencing>). The method is also used for animal [47] and plant studies [48].

8 Future Prospects

Since the completion of the human genome-sequencing project, there has been a focus on developing rapid new methodologies that will enable routine genomics studies and molecular testing in clinical settings. The Pyrosequencing technology was the first alternative sequencing method on the market. The technology is unique as it can be used in both a low- and a high-throughput format. The technology is the first next generation sequencing system allowing massively parallel clonal sequencing. The introduction of the technology has fundamentally altered genomic

research and allowed investigators to conduct experiments that were previously not feasible or affordable. For example, in gene-expression studies microarrays are now being replaced by sequencing-based methods. Today, several high-throughput sequencing strategies are available and the choice of methodology is dependent on scientific application, the cost, and the throughput. The Pyrosequencing technology has, due to its long- and high-quality reads, big potential to have a major impact on immunogenetics. It is especially suitable for human leukocyte antigen (HLA) typing [49]. The technology is also suitable for de novo sequencing applications, such as for detection of unknown organisms and viruses in human disease [50].

The low-throughput format of Pyrosequencing is now established as a very reliable method for diagnostic sequencing of short DNA stretches. The information it provides about sequence context for SNP analysis makes it the most robust method available at this time. An especially interesting field for Pyrosequencing technology is epigenetics where the detection of methylation and quantitative analysis of genomic methylation are of interest. The possibility for robust quantitative analysis is of interest also for forensic analysis [42–44] (*see* Chapters 26 and 27) and phylogenetic studies [51]. Pyrosequencing is also suitable for bacterial analysis of whole blood in sepsis [52].

The simplicity of the Pyrosequencing technology opens for improvements of both mechanical and chemical aspects. An example of mechanical improvement is the recent development of a smaller and cheaper Pyrosequencing sequencing system suitable for field studies [53]. Increasing read length is among the greatest challenges related to the chemistry of the Pyrosequencing method. The Pyrosequencing method relies on the cooperation of four different enzymatic reactions. One way to aid understanding of this complicated process is to use mathematical models. We have constructed a mathematical model of the Pyrosequencing reaction system utilizing the assumptions of irreversible Michaelis-Menten rate equations [54]. Both from the model and from practical experience we have identified factors important for improved performance of the chemical reactions. An especially interesting observation is that the high-throughput system allows sequencing over 1,000 bases, whereas the low-throughput system is limited to 100 bases. The two systems use different enzymes, as well as different procedures, for sequencing DNA. The incorporation efficiency is improved by more processive DNA polymerases and by longer incubation with nucleotides. A more processive DNA polymerase and a higher nucleotide concentration can dramatically improve the incorporation efficiency for the low-throughput system. A factor that is less obvious to improve is the observed false incorporation. This is clearly observed only for the low-throughput system. One possibility explanation for this phenomenon is impure

nucleotides. It is well known that commercial nucleotides contain impurities of different types. Impurities that affect the low-throughput system only are nucleotide polyphosphates and dinucleotide polyphosphates. Normally a nucleotide is contaminated with low amount of polyphosphates of the same base. In the high-throughput format these contaminants are washed away between each nucleotide addition, whereas in the low-throughput format they are accumulated, as they are not hydrolyzed by the apyrase. Although, present in low concentration, they have the potential to be used as substrate by the DNA polymerase and thereby, create a false incorporation. This type of undesirable “false-incorporation” will accumulate during time and thereby, limit the obtained read length. Contaminating enzymatic activities that convert one nucleotide into a new nucleotide by transferring phosphates between the different molecules during the sequencing process can also contribute to “false-incorporation” events. We believe that there is great potential for increasing Pyrosequencing’s read length, thereby extending the method’s applicability in new and different fields.

References

1. Nyrén P (2001) Method for sequencing DNA based on the detection of the release of pyrophosphate and enzymatic nucleotide degradation. Patents: US 6 258 568BI and WO98/28440
2. Ronaghi M, Uhlén M, Nyrén P (1998) A sequencing method based on real-time pyrophosphate detection. *Science* 281:363–365
3. Margulies M, Egholm M, Altman WE et al (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380
4. Runswick MJ, Powell SJ, Nyrén P et al (1987) Sequence of the bovine mitochondrial phosphate carrier protein: structural relationship to ADP/ATP translocase and the brown fat mitochondria uncoupling protein. *EMBO J* 6:1367–1373
5. Nyrén P, Lundin A (1985) Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. *Anal Biochem* 151:504–509
6. Nyrén P, Nore BF, Baltscheffsky M (1986) Studies on photosynthetic inorganic pyrophosphate formation in *Rhodospirillum rubrum* chromatophores. *Biochim Biophys Acta* 851:276–282
7. Nyrén P, Nore BF, Baltscheffsky M (1986) Inorganic pyrophosphate synthesis after a short light flash in chromatophores from *Rhodospirillum rubrum*. *Photobiophys* 11:189–196
8. Nyrén P (1987) Enzymatic method for continuous monitoring of DNA-polymerase activity. *Anal Biochem* 167:235–238
9. Melamede RJ (1985) Automatable process for sequencing nucleotide. US Patent 4863849
10. Ståhl S, Hultman T, Moks T et al (1988) Solid phase DNA sequencing using the biotin-avidin system. *Nucleic Acids Res* 16:3025–3038
11. Nyrén P (1994) Apyrase immobilized on paramagnetic beads used to improve detection limits in bioluminometric ATP monitoring. *J Biolumin Chemilumin* 9:29–34
12. Nyrén P, Pettersson B, Uhlén M (1993) Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay. *Anal Biochem* 208:171–175
13. Nyrén P, Karamohamed S, Ronaghi M (1997) Detection of single-base changes using a bioluminometric primer extension assay. *Anal Biochem* 244:367–373
14. Ronaghi M, Karamohamed S, Pettersson B et al (1996) Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem* 242:84–89
15. Nyrén P (1994) A method for the detection of cells, cell lysis or cell-lysing activity. Swedish patent application
16. Nyrén P, Edwin V (1994) Inorganic pyrophosphate-based detections: detection and enumeration of cells. *Anal Biochem* 220:39–45

17. Nyrén P, Edwin V (1994) Inorganic pyrophosphate-based detections: detection and quantification of cell lysis and cell-lysing activity. *Anal Biochem* 220:46–52
18. Karamohamed S, Nordström T, Nyrén P (1999) A real-time bioluminometric method for detection of nucleoside diphosphate kinase activity. *Biotechniques* 26:728–734
19. Karamouhamed S, Nilsson J, Nourizad K et al (1999) Production, purification, and real-time functional analysis of recombinant *Saccharomyces cerevisiae* MET3 adenosine triphosphate sulfurylase expressed in *Escherichia coli*. *Protein Expr Purif* 15:381–388
20. Ronaghi M (2000) Improved performance of pyrosequencing using single-stranded DNA-binding protein. *Anal Biochem* 286:282–288
21. Nordström T, Gharizadeh B, Pourmand N et al (2001) Method enabling fast partial sequencing of cDNA clones. *Anal Biochem* 292:266–271
22. Nordström T, Ronaghi M, Nyrén P (1999) Automation of a novel DNA sequencing method. In: Rod A, Pazzagli M, Kricka LJ, Stanley PE (eds) *Bioluminescence and chemiluminescence: perspective for the 21st Century*. John Wiley, Hoboken, NJ, pp 528–531
23. Gharizadeh B, Nordström T, Ahmadian A et al (2002) Long read pyrosequencing using pure 2'-deoxyadenosine-5'-O'-(1-thiotriphosphate) Sp-isomer. *Anal Biochem* 301:82–90
24. Eriksson J, Gharizadeh B, Nourizad N et al (2004) 7-deaza-2'-deoxyadenosine-5'-triphosphate as an alternative nucleotide for the pyrosequencing technology. *Nucleosides Nucleotides Nucleic Acids* 23:1583–1594
25. Gharizadeh B, Eriksson J, Nourizad N et al (2004) Improvements in pyrosequencing technology by employing sequenase polymerase. *Anal Biochem* 330:272–280
26. Eriksson J, Nordström T, Nyrén P (2003) Method enabling firefly luciferase based bioluminometric assays at elevated temperature. *Anal Biochem* 314:158–161
27. Eriksson J, Gharizadeh B, Nordström T et al (2004) Pyrosequencing technology at elevated temperature. *Electrophoresis* 25:20–27
28. Nordström T, Ronaghi M, Morgenstern R et al (2000) Direct analysis of single nucleotide polymorphism on double-stranded DNA. *Biotechnol Appl Biochem* 31:107–112
29. Nordström T, Nourizad K, Ronaghi M et al (2000) Method enabling pyrosequencing on double-stranded DNA. *Anal Biochem* 282:186–193
30. Nordström T, Alderborn A, Nyrén P (2002) Method for one-step preparation of double-stranded DNA template applicable for use with pyrosequencing technology. *J Biochem Biophys Methods* 52:71–82
31. Garcia CA, Ahmadian A, Garizadeh B et al (2000) Mutation detection by pyrosequencing: sequencing of exons 5 to 8 of the p53 tumor suppressor gene. *Gene* 253:249–257
32. Nourizad N, Gharizadeh B, Nyrén P (2003) Method for clone checking. *Electrophoresis* 24:1712–1715
33. Gharizadeh B, Ghaderi M, Donnelly D et al (2003) Multiple-primer DNA sequencing method. *Electrophoresis* 24:1145–1151
34. Gharizadeh B, Ohlin A, Mölling P et al (2003) Multiple group-specific sequencing primers for reliable and rapid DNA sequencing. *Mol Cell Probes* 17:203–210
35. Gharizadeh B, Oggionni M, Zheng B et al (2005) Type-specific multiple sequencing primers: a novel strategy for reliable and rapid genotyping of human papillomaviruses by pyrosequencing technology. *J Mol Diagn* 7:198–205
36. Leamon JH, Lee WL, Tartaro KR et al (2003) A massively parallel PicoTiterPlate based platform for discrete picoliter-scale polymerase chain reaction. *Electrophoresis* 24:3769–3777
37. Ahmadian A, Gharizadeh B, Gustafsson A et al (2000) Single nucleotide polymorphism analysis by pyrosequencing. *Anal Biochem* 280:103–110
38. Gruber JD, Colligan PB, Wolford JK (2002) Estimation of single nucleotide polymorphism allele frequency in DNA pools by using Pyrosequencing. *Hum Genet* 110:395–401
39. Uhlmann K, Brinckmann A, Toliat MR et al (2002) Evaluation of a potential epigenetic biomarker by quantitative methyl-single nucleotide polymorphism analysis. *Electrophoresis* 23:4072–4079
40. Ahmadian A, Lundeberg J, Nyrén P et al (2000) Analysis of the p53 tumor suppressor gene by pyrosequencing. *Biotechniques* 28:140–147
41. Goriely A, McVean GA, Rojmyr M et al (2003) Evidence for selective advantage of pathogenic FGFR2 mutations in the male germ line. *Science* 301:643–646
42. Andreasson H, Asp A, Alderborn A et al (2002) Mitochondrial sequence analysis for forensic identification using pyrosequencing technology. *Biotechniques* 32:124–133
43. Allen M, Andreasson H (2005) Mitochondrial D-loop and coding sequence analysis using pyrosequencing. *Methods Mol Biol* 297:179–196

44. Cebula TA, Brown EW, Jackson SA et al (2005) Molecular applications for identifying microbial pathogens in the post-9/11 era. *Expert Rev Mol Diagn* 5:431–445
45. Gharizadeh B, Norberg E, Löffler J et al (2004) Identification of medically important fungi by pyrosequencing technology. *Mycoses* 47:29–33
46. Gharizadeh B, Kalantari M, Garcia C et al (2001) Typing of human papillomavirus (HPV) by pyrosequencing. *Lab Invest* 81:673–679
47. Milan D, Jeon JT, Looft C et al (2000) A mutation in PRKAG3 associated with excess glycogen content in pig skeletal muscle. *Science* 19:1248–1251
48. Mochida K, Yamazaki Y, Ogihara Y (2003) Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. *Mol Genet Genomics* 270:371–377
49. Danzer M, Niklas N, Stabentheiner S et al (2013) Rapid, scalable and highly automated HLA genotyping using next-generation sequencing: a transition from research to diagnostics. *BMC Genomics* 14:221
50. Feng H, Shuda M, Chang Y et al (2008) Clonal integration of a polyomavirus in human merkel cell carcinoma. *Science* 319:1096–1100
51. Hahn KR, Janzen TW, Thomas MC et al (2014) Single nucleotide repeat analysis of *B. anthracis* isolates in Canada through comparison of pyrosequencing and Sanger sequencing. *Vet Microbiol* 169:228–232
52. McCann CD, Jordan JA (2014) Evaluation of MolYsis™ Complete DNA extraction method for detecting *Staphylococcus aureus* DNA from whole blood in a sepsis model. *J Microbiol Methods* 99:1–7
53. Song Q, Wei G, Zhou G (2014) Analysis of genetically modified organisms by pyrosequencing on a portable photodiode-based bioluminescence sequencer. *Food Chem* 154:78–83
54. Svantesson S, Westermark PO, Hellgren-Kotaleski J et al (2004) A mathematical model of the pyrosequencing reaction system. *Biophys Chem* 110:129–145

Free ebooks ==> www.Ebook777.com

Chapter 2

PyroMark® Instruments, Chemistry, and Software for Pyrosequencing® Analysis

Martin Kreutz, Gerald Schock, Julia Kaiser,
Norbert Hochstein, and Ralf Peist

Abstract

Since the early 2000s, Pyrosequencing® technology has been adapted for various instrument platforms to enable users to examine the role of epigenetic DNA methylation in gene expression regulation, genetic markers for specific phenotypes in livestock, drug resistance development in pathogens, and polymorphisms in forensic samples of mitochondrial DNA.

The instruments, software, and chemistry have been modified to facilitate different sample throughputs and sample amounts. Just recently, major changes have been implemented to enable increased read length and more precise Pyrosequencing results. These improvements were made possible through a number of changes to various system components. In addition, assay development has been streamlined through the availability of optimized PCR and Pyrosequencing reagents, automated assay design tools, and a number of predesigned Pyrosequencing assays.

In future, instruments with smaller footprints and the ability to automate crucial steps of the Pyrosequencing protocol will be available and will provide even more convenient and standardized Pyrosequencing analysis with flexible throughput.

Key words Pyrosequencing®, PyroMark® instruments, Workflow, Data analysis

1 Introduction

Pyrosequencing® technology, which is based on the principle of sequencing-by-synthesis, provides quantitative data in a sequence context within minutes. Pyrosequencing enables researchers to accurately and quantitatively analyze DNA and RNA sequences. The PyroMark® instruments are fully integrated systems that provide real-time sequence information and are suitable for epigenetic and genetic analyses.

A sequencing primer is hybridized to a single-stranded, PCR-amplified DNA template (*see* Fig. 1). This DNA template is incubated with enzymes and substrates. With each dispensation, one of the four nucleotides is added to the reaction. If the dispensed

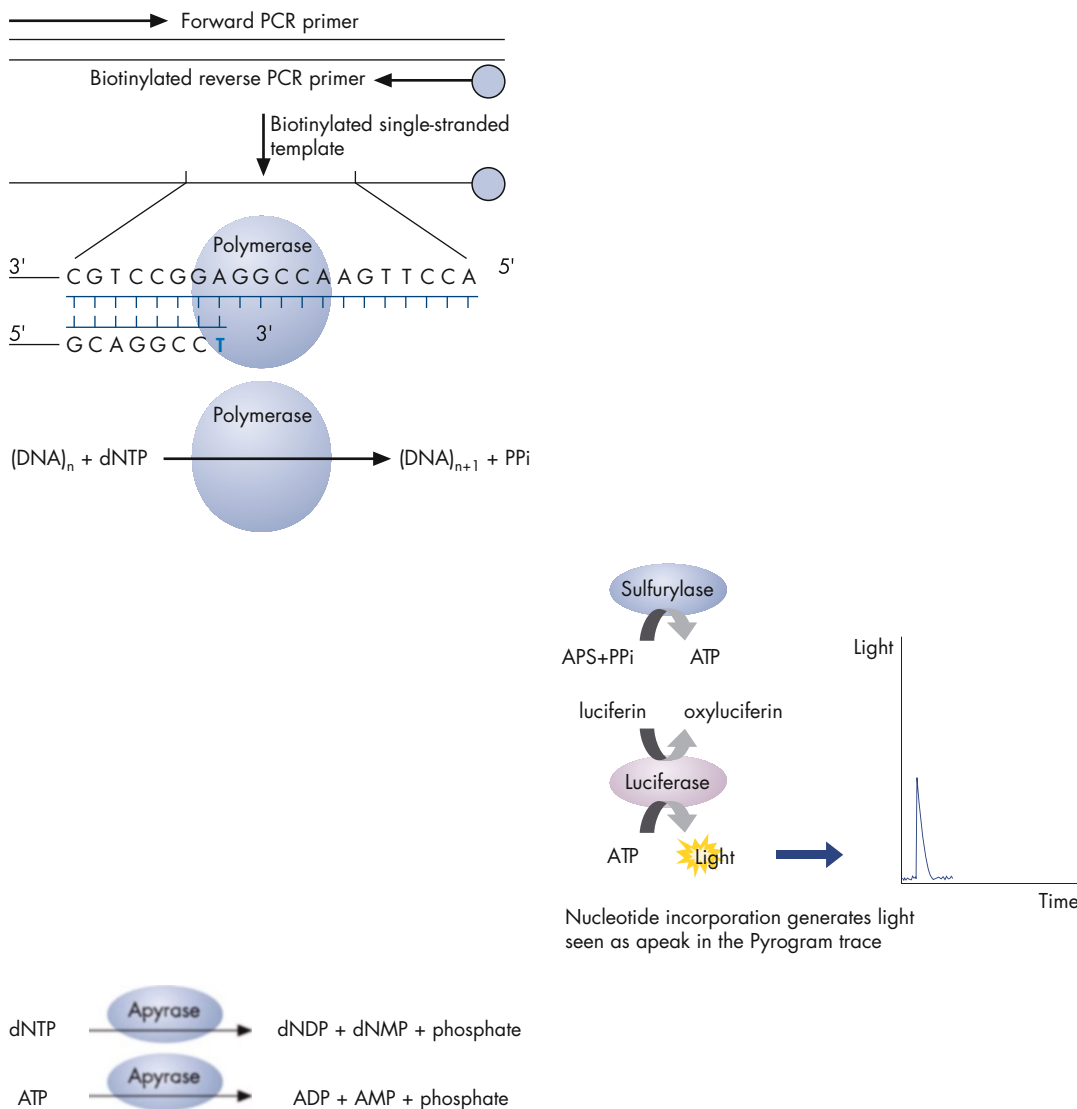


Fig. 1 Schematic illustration of the Pyrosequencing reaction cascade. After successful incorporation of a nucleotide using a single-stranded PCR fragment as template, the released PPi is converted to light by an enzyme cascade. The amount of light is proportional to the amount of incorporated nucleotides and is quantified by a CCD. Sequential addition of nucleotides allows quantitative decoding of the sequence to analyze. Apyrase degrades any unincorporated nucleotides before addition of the next nucleotide

nucleotide is complementary to the base in the template strand, it will be incorporated into the DNA strand by the DNA polymerase. This incorporation is accompanied by the release of pyrophosphate (PPi) in an amount equimolar to the amount of nucleotide incorporated [1].

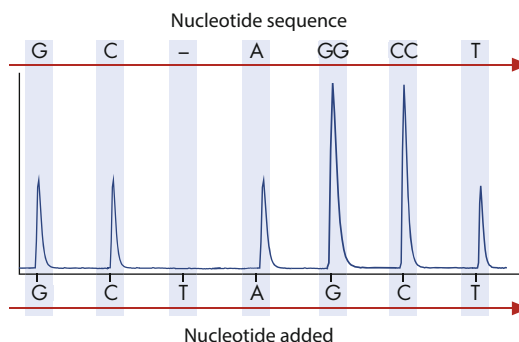


Fig. 2 An example of a Pyrogram showing a typical result from a Pyrosequencing assay. Two single peaks are detected, followed by a (control) T dispensation (no incorporation). The higher peaks for G and C after the A dispensation indicate that two homodimers follow a single A in the sequence. The detected sequence is GCAGGCCT

In the presence of adenosine 5'-phosphosulfate, the released PPi is quantitatively converted to ATP by the ATP sulfurylase. The ATP can be used by luciferase to convert luciferin to oxyluciferin, generating visible light in amounts proportional to the amount of ATP [2]. The emitted light is detected using charge-coupled devices (CCDs) and is displayed as a peak in the Pyrogram®.

Unincorporated nucleotides and ATP are continuously degraded by a nucleotide-degrading enzyme called apyrase. Apyrase completely degrades remaining nucleotides before the next nucleotide is added.

Nucleotides are added to the reaction one at a time. Since natural deoxyadenosine triphosphate (dATP) would be recognized by the luciferase, an alternative nucleotide, deoxyadenosine alpha-thio triphosphate (dATP α S), is used instead. It can be used efficiently by the DNA polymerase but is not recognized by the luciferase. With each added nucleotide, the complementary sequence is built up and the nucleotide sequence is determined from the peak pattern in the Pyrogram (Fig. 2).

2 Workflow

Prior to the Pyrosequencing reaction, a biotinylated PCR product must be generated. Because a single-stranded DNA template is required for Pyrosequencing, the second strand must be removed. To achieve this, the biotinylated PCR product is bound to streptavidin-coated Sepharose® beads, and the beads are captured with the vacuum tool on the vacuum workstation (Fig. 3), where they are thoroughly washed and subsequently denatured to



Fig. 3 The PyroMark Vacuum Workstation. The troughs are filled with buffers and reagents necessary for DNA single-strand generation. The DNA template, bound to Sepharose beads, is moved through the different troughs using the vacuum tool

generate single-stranded DNA suitable for Pyrosequencing. This template DNA is released into a reaction plate already containing the sequencing primer and reaction buffer. With the PyroMark Q48 Autoprep, the latest Pyrosequencing instrument which is still under development as of spring 2014, this manual step will be automated within the Pyrosequencing workflow.

Annealing of the sequencing primer to the target region on the single-stranded DNA template is achieved by heating the sample to 80 °C, followed by a cooling step. After annealing, the assay plate containing the samples is placed inside the instrument. A cartridge is filled with Pyrosequencing reagents and placed in the dispensing unit. The instrument is either started via a computer connected to the instrument (PyroMark Q96 ID and PyroMark Q96 MD), or a run file containing all data necessary for the run is transferred to the instrument using the USB port at the front of the instrument (PyroMark Q24). After the run is started by the user, the dispensing unit pressure, mixer speed, and temperatures of the heating block, process chamber lid, and coolant liquid are adjusted to preset levels. To ensure that the dispensation capillaries are flushed and filled with solution, enzyme and substrate mixtures are dispensed into the plate's priming well. Enzyme mixture and then substrate mixture are dispensed into all wells that will be used for the run. The pressure in the dispensing unit is increased. Nucleotides are dispensed into the priming well before being dispensed into the reaction wells. Nucleotides are added in a predefined order, with a 65 s lapse between each

nucleotide addition, to ensure that all enzymatic reactions are complete and that all unincorporated nucleotides have been degraded before the next nucleotide is added.

3 The Instruments

With several different Pyrosequencing instruments available, the right choice of instrument greatly depends on the researchers' needs (Fig. 4). The PyroMark Q24 combines ease-of-use, analysis versatility, and superior detection sensitivity. With the PyroMark Q24 MDx (not available in all countries), a dedicated instrument supports QIAGEN's CE-marked mutation and methylation analysis solutions using Pyrosequencing technology, in compliance with the European In Vitro Diagnostic Directive 98/79/EC. The PyroMark Q24 Advanced further plays on the strength of the PyroMark Q24 instrument with its improved functionality, longer read lengths, and additional run options. The PyroMark Q96 ID accommodates the need for a higher throughput by processing up to 96 samples in a single run.

Regardless of the instrument, all of the PyroMark systems are designed to make the Pyrosequencing workflow straightforward and efficient. Each step, from assay design to PCR amplification and preparation of sequencing templates, is supported by software, kits, reagents, and sample prep instrumentation optimized for Pyrosequencing.



Fig. 4 The PyroMark instruments. From *left to right*: PyroMark Q96 ID and PyroMark Q24. In addition, both PyroMark Q24 Advanced and PyroMark Q24 MDx are based on the PyroMark Q24 platform

4 PyroMark Q96 ID

The PyroMark Q96 ID was the first instrument developed to use the Pyrosequencing technology. The PyroMark Q96 ID originally used an Oracle database for managing the Pyrosequencing data, but now provides the alternative option to use the file-based PyroMark Q96 ID software version 2.5, which is similar to the software used by the PyroMark Q24 instrument.

The PyroMark Q96 ID is a powerful sequencing and quantification platform highly suited for use in epigenetics, mutation gene expression analysis, and microbial identification and resistance typing. With its 96-well format, automatic base-calling function, and dedicated software for methylation analysis, the PyroMark Q96 ID can handle any research question requiring true sequence information and quantification of genetic or epigenetic variation.

The PyroMark Q96 ID manages all steps necessary to rapidly sequence and analyze up to 96 samples. Reagents and nucleotides are dispensed to each well, and the emitted light signals are detected with sensitive CCD sensors. The data obtained can then be analyzed using the PyroMark software. The versatile and highly reliable PyroMark Q96 ID enables sequence-based quantification and detection of SNPs, insertion and deletion mutations, CpG sites, and generation of de novo sequence information.

To increase ease-of-use for researchers using the PyroMark Q96 ID, a new and even more user-friendly software is available. The PyroMark Q96 ID software version 2.5 enables comprehensive analysis of the obtained Pyrosequencing data. The updated software contains four analysis modes: AQ (allele quantification), SNP (analysis of SNPs and InDels), CpG (methylation analysis), and SQA (sequence identification). These four different types of analyses can be performed on the same plate, in the same run.

5 PyroMark Q24 and PyroMark Q24 MDx

Smaller in size than the PyroMark Q96 ID, the PyroMark Q24 instrument manages all steps necessary to rapidly analyze up to 24 samples. In contrast to the PyroMark Q96 instrument, the emitted light signals are detected by 24 CCD sensors—one sensor per well—in the PyroMark Q24, which greatly reduces signal crossover.

The PyroMark Q24 uses Pyrosequencing technology for real-time, sequence-based detection and quantification of sequence variants and epigenetic methylation. This platform is highly suited for analyzing CpG methylation, SNPs, mutations, insertion/deletions, STRs, variable gene copy number, as well as for microbial identification and resistance typing.

The PyroMark Q24 is designed as a stand-alone instrument. The data obtained is stored on the instrument hard drive and can be viewed on the instrument screen during a run. The completed run file can be transferred from the instrument using a USB stick, allowing data analysis on any computer with PyroMark Q24 software installed.

The PyroMark Q24 Software 2.0, installed on a PC, enables comprehensive analysis of the obtained results. The software contains three analysis modes: SQA (sequence analysis, de novo sequencing), CpG (DNA methylation analysis), and AQ (allele quantification). All modes can be used to analyze samples on the same plate, enabling different types of samples to be run at the same time. The AQ mode can be used for analyzing single- and multivariable positions, as well as di-, tri-, and tetra-allelic mutations. The CpG mode enables analysis of multiple consecutive CpG sites and provides a built-in control for the bisulfite treatment. The third mode, SQA, allows analysis of results for de novo sequencing, which is used in microbial identification, for example.

In compliance with European In Vitro Diagnostic Directive 98/79/EC, the PyroMark Q24 MDx (not available in all countries) is the dedicated PyroMark instrument for QIAGEN's CE-marked mutation and methylation analysis. The PyroMark Q24 MDx uses proven Pyrosequencing technology for real-time, sequence-based detection and quantification for in vitro diagnostic use in Europe. Furthermore, it is highly suited for genotyping mutations, evaluating disease-related DNA methylation patterns, validating biomarkers, and other assays in biomedical research.

6 PyroMark Q24 Advanced

Although all Pyrosequencing instruments are highly accurate and allow exact quantification, the reliable read length of the systems was previously fairly limited. Several factors contribute to this limitation. In the case of long stretches of homopolymers, not all nucleotides can be incorporated in a single dispensation. In most Pyrosequencing applications, countermeasures such as additional dispensations of the same nucleotide can be used to minimize this effect. However, if the sequence that is to be analyzed is unknown, as is the case for de novo sequencing, this can negatively affect the length that can reliably read or analyzed.

Depending on the particular assay, the annealing process can sometimes require a longer period of time, to ensure that all DNA templates have bound to a sequencing primer. If some templates remain unbound at the start of the reaction, an effect called “restart” can occur. If this happens, the subset of DNA templates that bind a sequencing primer after the reaction has already started will be sequenced with a delay, i.e., at a later nucleotide dispensation.

This results in an increase in the background noise of the Pyrosequencing reaction. Reduced light signal and an increase of background peaks due to enzyme exhaustion at the end of the Pyrosequencing reaction also limit the read length.

Since the PyroMark software has a thorough set of controls and rigorously checks that all performance criteria are met, it will recognize whenever a deviation from the expected peak pattern occurs. In such a case, the software will flag the result and display an appropriate warning message to ensure the quality of the acquired data.

To further improve the reliable read length and efficiency of analysis, the PyroMark Q24 Advanced system was developed. New knowledge along with thorough testing of how buffer components affect the Pyrosequencing reaction were used to improve the existing Pyrosequencing chemistry. These improvements include changes in the reaction buffer composition, as well as optimizations to the enzymes used. In addition, the instrument operation algorithms were updated to include the knowledge acquired from years of experience with the Pyrosequencing technology.

These updates to the chemistry and instrument algorithms in the PyroMark Q24 Advanced system result in significant reduction in background peaks, thereby increasing read length and reliability of the analysis later in the sequence. On average, read length of Pyrosequencing assays could be doubled by introducing these changes. For de novo sequencing reactions (SEQ mode), highly accurate read lengths of 140 or more bases can be obtained, depending on the sequence to be analyzed.

To reduce the probability of “restart” effects, changes were also made to the annealing procedure. The buffer volume for the annealing step was reduced in order to increase the sequencing primer concentration during the annealing step. This facilitates larger volumes of enzymes and substrate without changing the final volume of the sequencing reaction. To ensure constant and comparable conditions during the annealing step, the annealing takes place inside the instrument after the initial heating step. The PyroMark Q24 Advanced instrument ensures constant mixing during this step to further improve the annealing efficiency.

To ensure that all bases are incorporated at homopolymer positions, multiple nucleotide shots were introduced at a given dispensation. This so-called multishot allows complete incorporation of the nucleotide at this position. Nucleotide concentrations were adapted accordingly in the PyroMark Q24 Advanced chemistry to avoid overloading.

DNA methylation analysis requires an additional step before Pyrosequencing: bisulfite conversion of the target DNA prior to the PCR reaction. This bisulfite conversion leads to frequent poly T stretches in the nucleotide sequence, and analysis of CpG sites directly after such T homopolymers can be challenging due to

uncertain quantification of the light signal at these sites. The efficiency of the bisulfite treatment has a direct influence on the reliability of the result. To ensure that the required quality of the template is met, the PyroMark Q24 Advanced automatically chooses suitable bisulfite treatment controls each time a dispensation order is created.

Even if the efficiency of the bisulfite conversion meets the required criteria, a subset of cytosines might still be unconverted. If this occurs in a long T homopolymer, all of the thymine residues following the unconverted cytosine residues will not be incorporated immediately. After the next dispensation of cytosine, the thymine nucleotides will be incorporated in the following T dispensation. These additional nucleotides will have a negative impact on both reference peaks and controls. In the PyroMark Q24 Advanced, such long homopolymers containing converted cytosine positions are addressed by a special “catch up” dispensation. In such a case, cytosine will be dispensed in the same dispensation as the thymine. This ensures that after the T dispensation, the entire homopolymer has been completed and no thymine remains unincorporated (Fig. 5). This “catch up” dispensation is omitted

Sequence to analyze:

TTGYGTTYGTGTTTGTGAYGTTAGGTTTTAGATTGGTGGTYGATTTTATTATATTAGGYGTTAYGTAYGTGTAGAAGATG
TTTATGTATAATTGAGATGYGYGGAGGAAAATTGTTTGTTAGGTAT

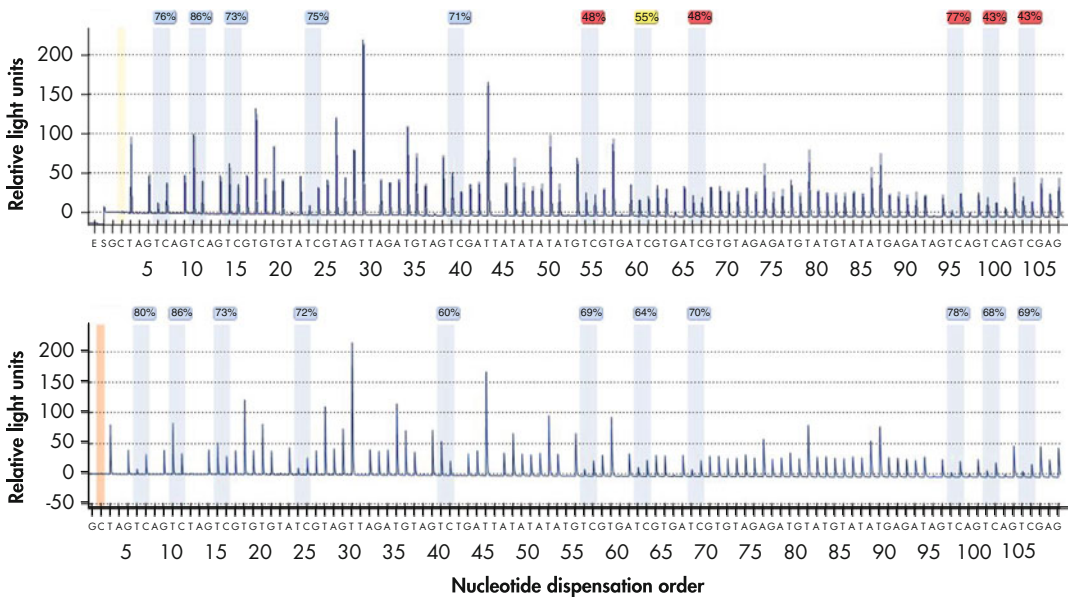


Fig. 5 DNA methylation analysis of the LOX gene locus. PyroMark Q24 (*upper panel*) and PyroMark Q24 Advanced (*lower panel*) were used to analyze 11 CpG sites in the LOX gene locus of highly methylated DNA using the same assay. PyroMark Q24 Advanced enables accurate analysis of all 11 CpG sites, due to its reduced background noise, which otherwise can lead to uncertain or invalid results at CpG sites located further from the sequencing primer

in case of a CpG site at the end of the homopolymer. Since T homopolymers occur frequently in CpG analysis, and since these stretches can easily be ten or more bases in length, the catch up of these positions allows a significant improvement of the quantification of cytosine methylation, especially later in the sequence.

DNA methylation usually occurs at CpG sites. Depending on the target organism however, other sites might show cytosine methylation as well. To enable analysis of methylation in non-CpG sites, a CpN mode was added to the PyroMark Q24 Advanced. In the CpN mode, the researcher can choose which of the cytosines present in the sequence are part of a potential variable position. Each potential variable position can then be selected for analysis. The dispensation order created by the PyroMark Q24 Advanced software will allow quantification at all the chosen positions. In case consecutive cytosine residues are chosen for methylation analysis, an average methylation frequency will be calculated.

The PyroMark Q24 Advanced software, installed on a PC, enables comprehensive analysis of Pyrosequencing results. The software contains four analysis modes—AQ, SNP, CpG, and SEQ. The AQ mode can be used for a variety of quantification studies of mutations such as SNPs and InDels. It is suitable for analyzing single- and multivariable positions, as well as di-, tri-, and tetra-allelic mutations. The SNP mode provides genotype analysis of SNPs and InDels. The CpG mode enables methylation analysis of single or multiple CpG or CpN sites and provides a built-in control for the bisulfite treatment. The SEQ mode is used for base-calling of unknown sequences.

The PyroMark Q24 Advanced also includes improved tools for run analysis and troubleshooting. If a problem occurs during the run, or if the system detects an inconsistency with an assay, the software provides specific warning information for each individual well. A “Warning Info” button gives access to additional information about the warning along with recommendations for troubleshooting and preventing its occurrence in subsequent assays.

Software packages to upgrade PyroMark Q24 systems to PyroMark Q24 Advanced are available.

7 PyroMark Q48 Autoprep

Still under development as of spring 2015, the PyroMark Q48 Autoprep is the newest Pyrosequencing instrument from QIAGEN. Using the same Pyrosequencing technology with a similar chemistry and software as the PyroMark Q24 Advanced, the PyroMark Q48 Autoprep offers a setup that no longer requires manual single-strand DNA generation and therefore streamlines the Pyrosequencing workflow. In addition to the reagents required for the Pyrosequencing reaction, buffers for generating the

single-stranded DNA are dispensed automatically by the instrument cartridges. The protocol starts with the transfer of biotinylated DNA templates into wells of a Q48 rotor disc and addition of magnetic streptavidin beads. Similar to the PyroMark Q24, a run file containing all necessary information is transferred to the instrument using a USB stick. Alternatively, data exchange between the PyroMark Q48 Autoprep software and the PyroMark Q48 Autoprep instrument can be done via a network connection. The highly intuitive touchscreen interface of the PyroMark Q48 Autoprep then guides the user through all steps necessary to prepare and start the Pyrosequencing run.

8 Pyrosequencing Assays

A large number of Pyrosequencing assays are described in the literature, and predesigned assays for selected targets are offered by QIAGEN. In addition, PyroMark CpG Assays, which are designed by a tailored design algorithm for high specificity and successful CpG methylation results, are the only genome-wide, predesigned methylation assays for Pyrosequencing analysis. QIAGEN's upgraded assay database includes over 84,000 individual assays for gene-specific human, mouse, and rat CpG sites. Predesigned assays for validating results from methylation arrays such as the HumanMethylation450K BeadChip array are also available.

Researchers preferring to design their own assays may use the PyroMark Assay Design Software 2.0 for assay design. After providing the target sequence along with surrounding sequence information, the Assay Design Software will generate suitable Pyrosequencing assays and rate the success chance of each generated assay, based on parameters such as primer length and positioning, GC content, potential stem loops, and other factors. The designed assays are optimized for use with all PyroMark instruments.

References

1. Nyrén P (1987) Enzymatic method for continuous monitoring of DNA-polymerase activity. *Anal Biochem* 167:235–238
2. Nyrén P, Lundin A (1985) Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. *Anal Biochem* 151:504–509

Free ebooks ==> www.Ebook777.com

Chapter 3

Software-Based Pyrogram® Evaluation

Guoli Chen, Matthew T. Olson, and James R. Eshleman

Abstract

Pyrosequencing® is a widely used technology to detect gene mutations in a molecular research or diagnostics laboratory. Compared to Sanger sequencing, it is inherently more quantitative with a superior limit of detection, although it has a shorter read length and has difficulty with homopolymeric sequences.

Results of Pyrosequencing experiments are typically presented as traces with sequential peaks, called Pyrograms®. For the majority of clinical diagnostic cases, Pyrograms are straightforward to read. However, there are occasionally complex results that are uninterpretable or difficult to interpret. In this chapter, we demonstrate a computer software, named Pyromaker that has been developed to help with the analysis of Pyrograms. Pyromaker is a freely and publically available software program to assist in the recognition of patterns of mutations, interpretation of difficult or ambiguous testing results and design of an optimal strategy to detect potential mutations by generating simulated Pyrograms. In addition to help diagnostic activities, Pyromaker can also be used as a virtual and user-friendly educational tool to teach newcomers the fundamental mechanism of Pyrosequencing, and correct interpretation of actual Pyrosequencing data.

Key words Pyrosequencing®, Pyrogram®, Pyromaker, Software

1 Introduction

The most commonly used sequencing technologies today in the molecular diagnostic laboratory are still Sanger sequencing and Pyrosequencing®, despite that next generation sequencing is likely to replace most of the current platforms in the future. Pyrosequencing is a “sequencing-by-synthesis” approach [1] that is used to detect DNA mutations in a well-defined region of a gene, and is also the basis for several next generation sequencing models [2]. In general it provides a limit of detection of approximately 5 % for mutant alleles, superior to Sanger sequencing (*see* also Chapter 4). Also, it is inherently more quantitative [3, 4]. The fundamental basis of Pyrosequencing is monitoring release of pyrophosphate in real time when each dNTP is sequentially dispensed into a DNA-synthesizing reaction. When the injected dNTP is complementary to the template and added into the DNA strand, a molecule of inorganic pyrophosphate is freed. Pyrophosphate is

used to produce ATP that is a coenzyme for the enzyme luciferase, which facilitates oxidization of luciferin to oxyluciferin, consequently light is generated and detected.

Pyrosequencing results are most often presented as a horizontal line with sequential peaks, so-called Pyrograms[®]. Presence or absence of peaks at sequential dNTP dispensation positions describes the sequencing information, and peak heights are proportional to the number of nucleotides incorporated. When a mutation is present, the pattern of peaks is changed compared to wild type. It could be disappearance of expected peaks, emergence of novel peaks, and/or changed peak heights. Commonly it is straightforward to read a Pyrosequencing result from a well-designed assay, but occasionally the data are ambiguous or hard to interpret, even for an experienced person. Moreover, due to its fundamental differences from the traditional sequencing method, Sanger sequencing, Pyrosequencing may be hard for beginners to quickly catch its working principles, precisely understand the relationship between sequence information and dispensation/peak sequences, and thereby correctly read the Pyrograms. Pyromaker provides a virtual tool to show all possible Pyrogram patterns for different mutations at a certain dispensation order for educational purposes. Different dispensation sequences can produce totally different Pyrograms or identical DNA sequences, so often it is critical to design an optimal dispensation sequence in order to get clear and distinctive identification of all possible sequence changes before Pyrosequencing is actually performed. To assist with all the aspects mentioned above, computer programs working directly or indirectly on actual Pyrograms are developed. Pyromaker is a great example and can be used for multiple purposes, as discussed in detail below.

2 Pyromaker

Pyromaker is a freely available and Web-based software program that uses an HTML/JavaScript client interface to collect user inputs. After the user input is entered, Pyromaker produces simulated traces for the Pyrosequencing results. Simulated Pyrograms can then be matched with original Pyrograms for various purposes [5, 6]. The software has been extensively validated with multiple genes carrying mutations commonly seen in cancers, and it is an open system and potentially works for many more genes suitable for Pyrosequencing to detect mutations.

As shown in Fig. 1a (<http://pyromaker.pathology.jhmi.edu>, Johns Hopkins University School of Medicine, Baltimore, Maryland), the interface of the software application is straightforward. The user inputs wild type and mutant sequences, status of being heterozygous or homozygous, and percentage of tumor cells.

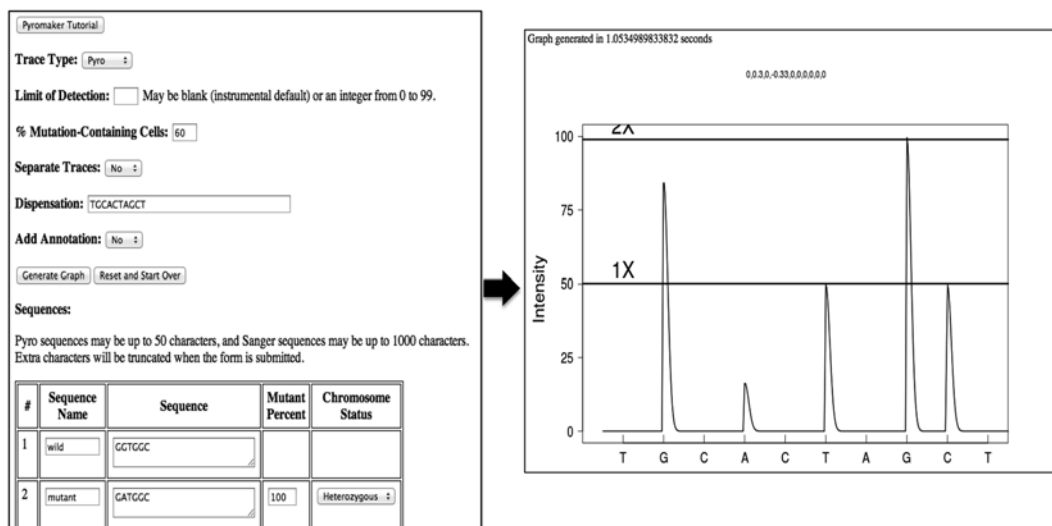


Fig. 1 A screenshot of the Pyromaker interface. After user-directed parameters are entered, a simulated Pyrogram is generated.

With the information, Pyromaker first calculates the percentage of mutant and wild-type alleles. The user can choose to use either a default or customized dispensation. Once all the information is submitted, the data are revalidated to ensure uniform sequence length, the inclusion of only A, C, G, and T characters in the sequence, and the presence of all parameters. If the inputs pass validation, a simulated Pyrogram is returned to the user in PDF format (Fig. 1b). The peak morphology is drawn as a linear increase to the maximum followed by decay phase as instructed by mimicking the Gaussian curve from the maximum point to four standard deviations (SD). This piecewise function and the SD are chosen based on visual comparisons of the measured Pyrograms against the virtual peak widths. A concise and clear tutorial instructing users to use the software and demonstrating basic concepts related to practice of Pyrosequencing is posted on the Pyromaker Web site (<http://pyromaker.pathology.jhmi.edu>).

Pyromaker can be helpful for several purposes:

2.1 Educational Function

Pyromaker can be used as a teaching tool to introduce newcomers to the basic concepts of the technique and elementary data interpretation skills. As briefly explained in the Web site tutorial, Harrington CT et al. [7] discussed how to employ the program to introduce the fundamental principles of Pyrosequencing, including (1) a peak is only produced when the correct nucleotide is injected and incorporated, (2) peak heights are proportional to the number of nucleotides incorporated, (3) different dispensation sequences may result in different Pyrograms for identical DNA sequences,

(4) suboptimal dispensation sequences can mask mutations, or produce remarkably complex Pyrograms.

For a person who is new to Pyrosequencing, it is also useful to recognize basic Pyrogram patterns for wild-type and commonly seen mutations of the genes tested in the laboratory. For example, Pyromaker is able to produce Pyrograms for all possible base substitution mutations in KRAS codon 13 (Fig. 2), and all mutations give distinct patterns.

After being familiar with the analysis of simple mutations, one can continue to generate simulated Pyrograms and gain some direct impression for more complicated mutations. Even further, one can learn how to design dispensation sequences to optimize the strategy to detect specific mutations.

2.2 Interpretation of Difficult Cases

Occasionally complicated mutations can lead to difficult Pyrogram patterns that cannot be definitively resolved solely based on the actual Pyrograms. Additional methodologies such as Sanger sequencing be helpful, but are usually time-consuming and cost-ineffective with a risk of retained ambiguity. More easily and quickly, simulated Pyrograms generated by Pyromaker can aid in the analysis of complex Pyrosequencing data. Here two real clinical cases are used to better explain the support in these situations. As shown in Fig. 3, Pyrosequencing of KRAS codons 12 and 13 from two cancer patients was originally performed, but the Pyrograms could not be clearly interpreted. Changes present in the Pyrogram of case 1 compared to the wild type (Fig. 3a) are complex: as numbered, extra T, C, and T peaks (Fig. 3b). There are also reduced peak heights at other positions (arrowheads). For case 2, a reduction of a G peak height, an extra A peak, a reduction of a T peak height, and an increased G peak height are noted (Fig. 3c).

Pyromaker can be used to resolve difficult Pyrosequencing results in one of the several ways: (1) hypothesis testing, (2) iterative refinement, and (3) matching against a look-up table. For the two cases discussed here, Sanger sequencing revealed mutations of two adjacent nucleotides in codon 12 for both cases, compared to the wild type. However, it still remained unclear whether the mutations are on the same allele or two different alleles. For case 1, in particular, it may due to a two-base mutation (GGT to TTT) on a single allele or a mixture of two single nucleotide mutations (TGT and GTT) on two separate alleles. Similarly, for case 2, it could be GGT to GAG on a single allele or a mixture of GAT and GGG from two alleles. These hypotheses were tested by generating virtual Pyrograms with Pyromaker and then matching the patterns to the actual data. As expected, the wild-type Pyrosequencing pattern was perfectly consistent between the virtual Pyrogram (Fig. 4a) and the actual one (Fig. 3a). In terms of the two hypotheses for the first case, one can easily tell that the Pyrosequencing pattern for TTT generated by the software qualitatively matched the actual

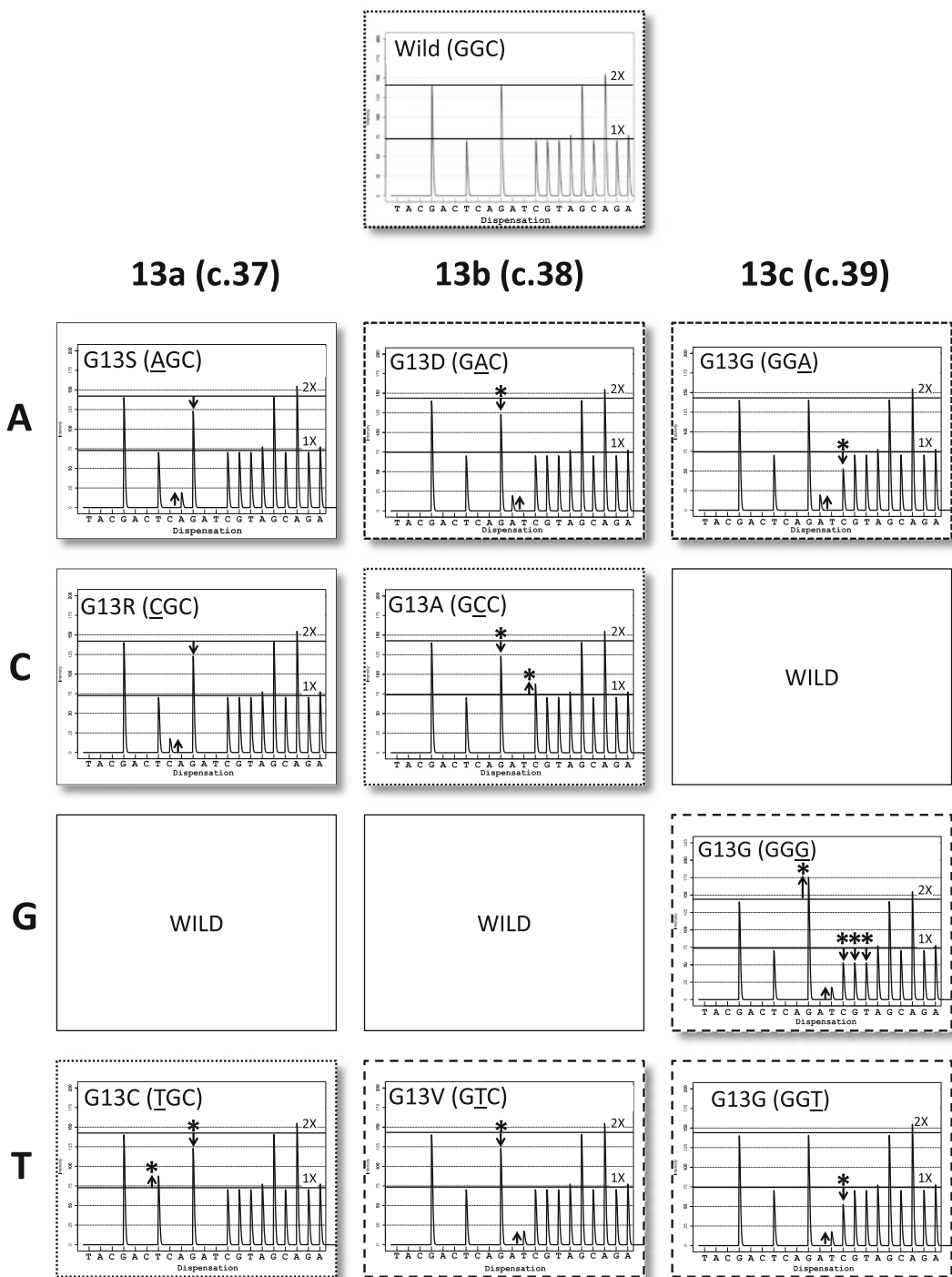


Fig. 2 Simulated Pyrograms generated by the computer program for all possible *KRAS* codon 13 mutations. Virtual Pyrograms for wild-type (GGC) and nine mutations are shown. Some mutations are qualitatively distinct from wild-type and all of the other mutations (AGC and CGC) whereas others are only quantitatively distinct (GAC vs. GGA, short dashed line borders), (wild-type vs. GCC vs. TGC, dotted borders), (GGG vs. GTC vs. GGT, long dashed line borders). Note that all three 13c mutations are silent. Reprinted from Chen G, Olson MT et al. (2012) A virtual Pyrogram generator to resolve complex Pyrosequencing results. *J Mol Diagn* 14(2): 149–159, with permission from Elsevier

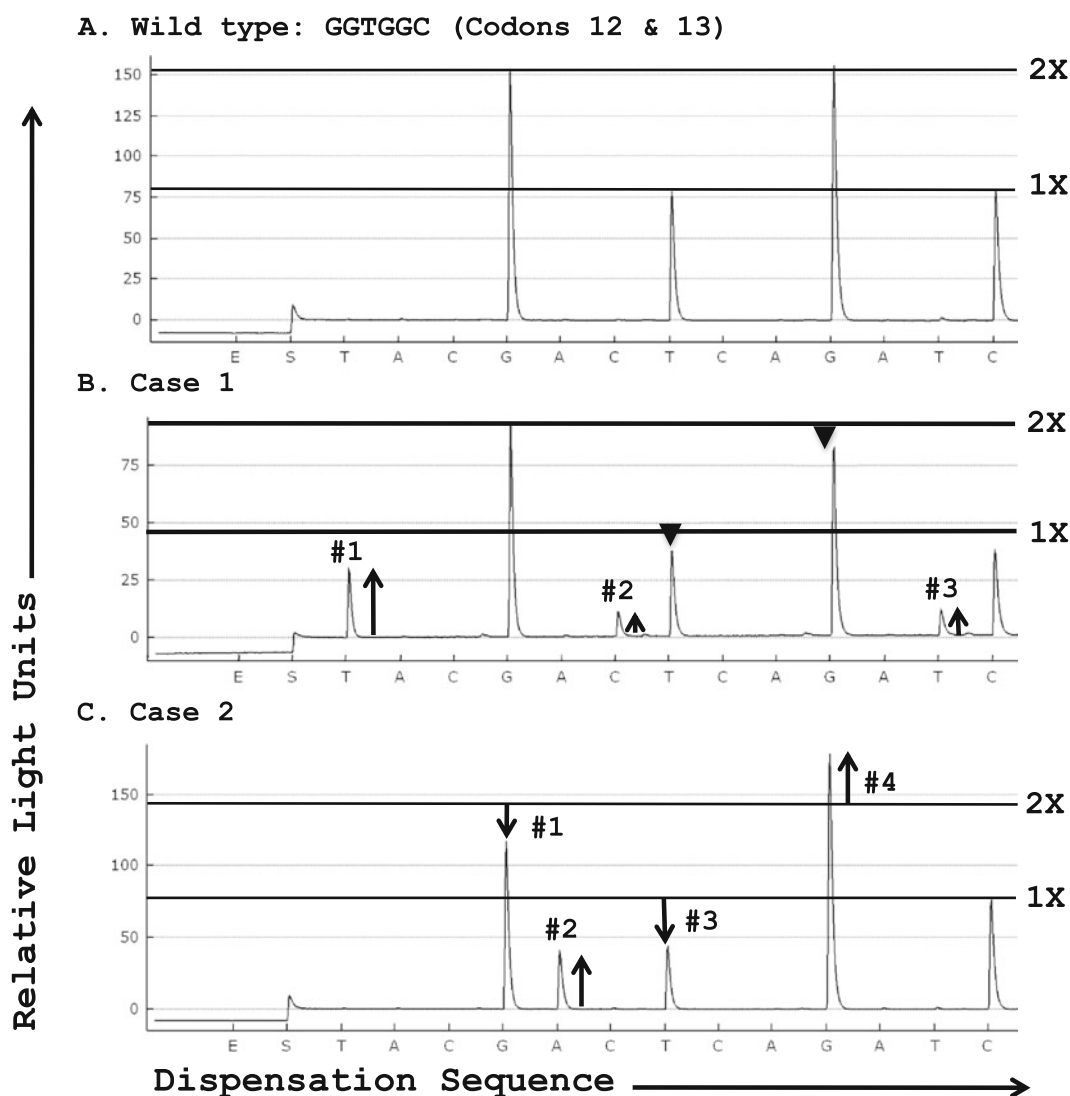


Fig. 3 Complex Pyrosequencing results for two clinical cases with KRAS mutations. (a) Pyrogram for wild-type KRAS; (b) Pyrogram for case 1; (c) Pyrogram for case 2. Reprinted from Chen G, Olson MT et al. (2012) A virtual Pyrogram generator to resolve complex Pyrosequencing results. *J Mol Diagn* 14(2): 149–159, with permission from Elsevier

clinical data. In contrast, the other hypothesis, TGT/GTT on two separate alleles (Fig. 4b, right panel) produced a clearly different pattern from Fig. 3b. Similarly, for case 2, the virtual Pyrogram for the GAG mutation (Fig. 4c, left) qualitatively matched the four characteristic changes generated by the mutation in the original Pyrogram (Fig. 3c), whereas a mixture of GAT/GGG alleles apparently led to a different pattern (Fig. 4c, right).

The second strategy is iterative mode working with Pyrosequencing results only. Therefore it can be done without

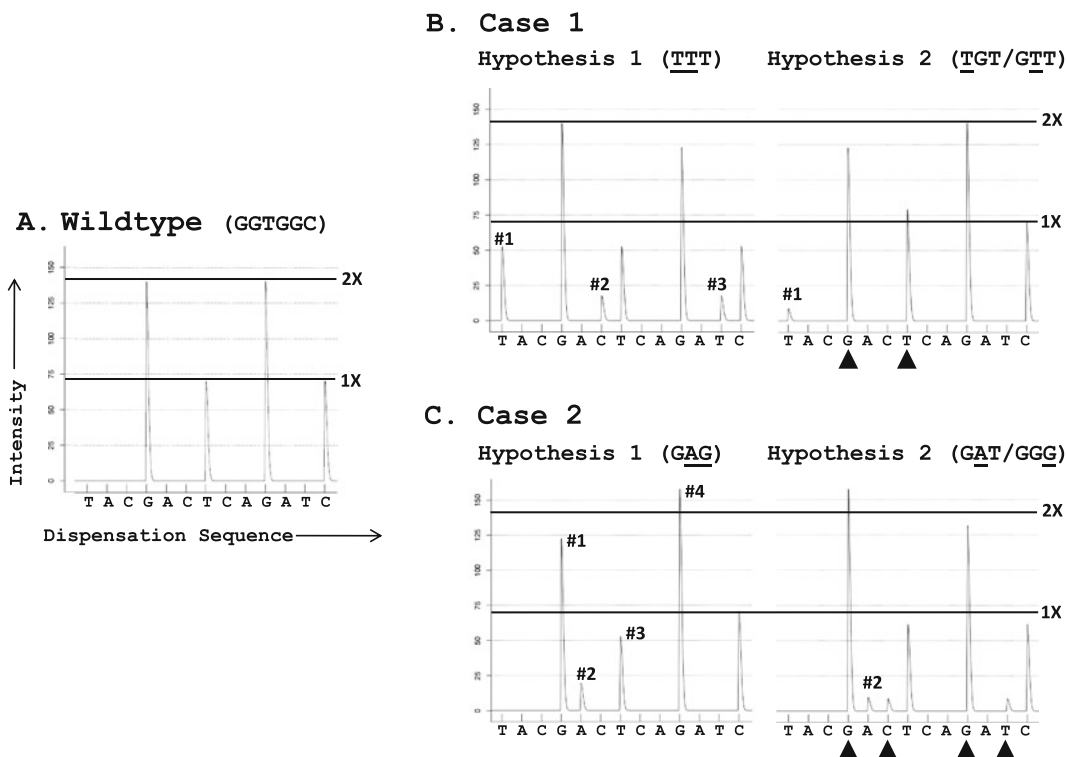


Fig. 4 Simulated Pyrograms generated by the computer program for hypotheses. (a) Simulated Pyrograms for wild-type KRAS; (b) Simulated Pyrograms for the hypotheses of case 1; (c) Simulated Pyrograms for the hypotheses of case 2. Reprinted from Reprinted from Chen G, Olson MT et al. (2012) A virtual Pyrogram generator to resolve complex Pyrosequencing results. *J Mol Diagn* 14(2): 149–159, with permission from Elsevier

using additional investigation to first generate hypotheses. Briefly, Pyromaker can be used in this way by trying all presumed mutated bases onto the wild-type Pyrogram to eventually reproduce a Pyrogram exactly matching the actual mutant one. As discussed, the Pyrogram for the first case contains three novel peaks (Fig. 3b). To obtain the first T peak (#1) ahead of the wild-type G, a T was added in the place of the first G of GGT (GGT>TGT) to generate a Pyrogram; however, this does not result in the complete changes but yields only the first extra T (Fig. 5a, far left). To proceed and obtain the C peak (#2 in Fig. 3b) ahead of the wild-type T, one must replace the G in the 12b position (TGT) with a second T, so ending up with TTT (Fig. 5a, middle left).

Alternatively, the T in the 12c position (TGT) can be replaced with a G to produce 4 Gs in a row (TGG/GGC, codons 12/13, Fig. 5a, middle right), thereby moving the T peak into the G position. Both sequences produce the three signature peaks shown in the clinical Pyrogram (Fig. 3b), but only the TTT Pyrogram yields the correct peak height ratios for the three novel peaks (3:1:1, respectively). A third possibility is TGC (Fig. 5a, far right), which

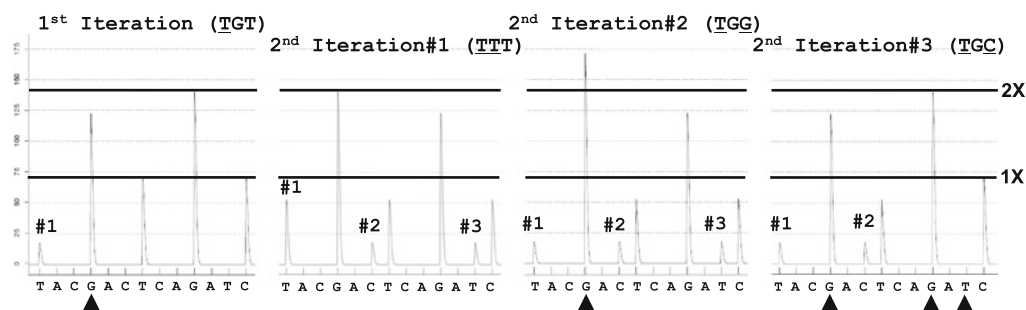
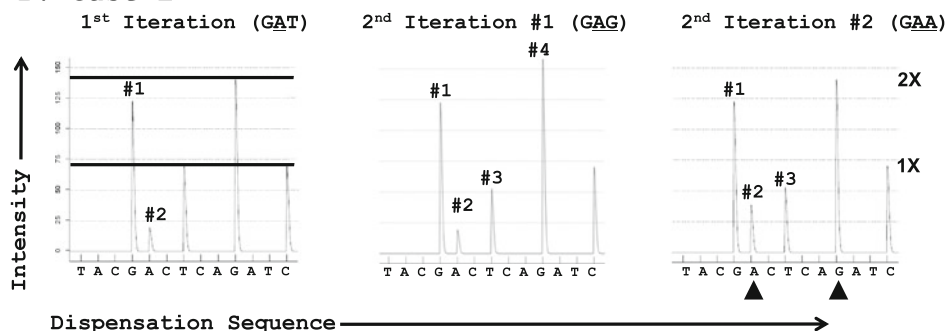
A. Case 1**B. Case 2**

Fig. 5 Simulated Pyrograms generated by the computer program in iterative mode. (a) interpretation of case 1 Pyrosequencing trace; (b) interpretation of case 2 Pyrosequencing trace. Reprinted from Chen G, Olson MT et al. (2012) A virtual Pyrogram generator to resolve complex Pyrosequencing results. *J Mol Diagn* 14(2): 149–159, with permission from Elsevier

however only provides two of three novel peaks. With a similarly strategy, one can also identify the underlying mutation for case 2 (Fig. 5b). Essentially all possibilities at each position can be tried (in order of the greatest likelihood) to identify the mutation with a perfectly matched pattern.

Hypothesis testing requires either additional sequencing methods or extensive experience with the gene and dispensation sequence so that an educated guess can be made. As such, it is most useful for laboratories that employ multiple sequencing strategies routinely and for simple Pyrograms, in which a mutation causes a small number of changes. Similarly, the iterative refinement process of methodically testing virtual Pyrograms and refining the mutant sequence to make a better match become factorially more complex with each increase in the dispensation sequence length and with the number of peak changes that result from the mutant sequence. Thus, these two strategies are limited to Pyrograms with a few (2 or 3) sequence changes. For more complex Pyrograms, it is more efficient to do an a priori calculation of all possible peak perturbations for every known mutation. These can be stored in a look-up

table. That way, the user needs only to note which peaks are altered and find the mutation with the same set of altered peaks in the look-up table. A confirmatory virtual Pyrogram can be generated to ensure an exact match. In addition, for peak sets that can be caused by multiple mutations, several Pyrograms may need to be generated in order to find the best match. This look-up table strategy was described for BRAF mutations, which tend to induce numerous peak alterations [6]. This look-up table approach is generally applicable to any gene with well-characterized mutation hot spots.

2.3 Optimization of Strategy for Mutation Detection

As previously mentioned, an optimal dispensation order is critical in a Pyrosequencing assay to clearly and correctly identify genetic variants as it intends to. To launch an assay designated to identify mutations in a certain region of a gene, therefore, optimization of the dispensation sequences is required before the actual Pyrosequencing can be performed. Pyromaker is a useful tool for this purpose because it allows one to confirm that a proposed dispensation detects all the potential mutations. It basically offers a platform where one can try all different possible dispensation orders till an ideal one or more are found. It is obviously much more convenient, quick, and economic than going thorough real wet lab Pyrosequencing steps. Similarly, for an existing Pyrosequencing assay, if new or unexpected genetic variants emerge to blur the current Pyrosequencing data, one can always use Pyromaker to modify the current strategy or generate a new one to accommodate new expectations.

In addition to design of dispensation orders, Pyromaker can be used to help with the optimization or simplification of mutation detection in other ways. The most efficient way to use Pyromaker for this purpose is to employ the look-up table and choose the dispensation sequence based on the smallest number of peak alterations, the clearest determination of the highest frequency mutations, and the fewest degenerate peak sets.

3 Summary

Pyromaker allows for user-oriented generation of virtual Pyrograms that can be compared with the actual data. This feature can be significantly helpful in a process of clarifying difficult or confusing Pyrograms by offering clues, confirming hypothesis, and/or adding more confidence. Although so far Pyromaker has been validated only with mutations in limited regions of several genes [5, 6], it is an open system so that one can easily test this web-based tool with other genes or DNA sequences of interest. Also, one does not even have to limit its application to Pyrosequencing-based analysis. Before conducting actual the

Pyrosequencing experiment to help with ambiguous data from other sequencing methodologies, all possibilities can be input into Pyromaker and pretested.

In addition to Pyromaker simulating Pyrograms, some software has been developed to directly evaluate original Pyrograms. In addition to the accompanying Pyrosequencing software from Qiagen, for example, a software has been used to automatically analyze Pyrograms for *EGFR*, *KRAS*, and *BRAF* mutations in parallel with manual analysis [8]. Also, another software tool was reported to be used for the evaluation of single nucleotide polymorphisms based on pattern recognition [9].

References

1. Nyren P (1987) Enzymatic method for continuous monitoring of DNA polymerase activity. *Anal Biochem* 167:235–238
2. Rothberg JM, Hinz W, Rearick TM et al (2011) An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475: 348–352
3. Dufort S, Richard MJ, De Fraipont F (2009) Pyrosequencing method to detect KRAS mutation in formalin-fixed and paraffin-embedded tumor tissues. *Anal Biochem* 391:166–168
4. Tsiatis AC, Norris-Kirby A, Rich RG et al (2010) Comparison of Sanger sequencing, Pyrosequencing, and melting curve analysis for the detection of KRAS mutations: diagnostic and clinical implications. *J Mol Diagn* 12:425–432
5. Chen G, Olson MT, O’neill A et al (2012) A virtual pyrogram generator to resolve complex Pyrosequencing results. *J Mol Diagn* 14: 149–159
6. Olson MT, Harrington C, Beierl K et al (2014) BRAF Pyrosequencing analysis aided by a lookup table. *Am J Clin Pathol* 141:639–647
7. Harrington CT, Lin EI, Olson MT et al (2013) Fundamentals of Pyrosequencing. *Arch Pathol Lab Med* 137:1296–1303
8. Shen S, Qin D (2012) Pyrosequencing data analysis software: a useful tool for EGFR, KRAS, and BRAF mutation analysis. *Diagn Pathol* 7:56
9. Ahmadian A, Gharizadeh B, Gustafsson AC et al (2000) Single-nucleotide polymorphism analysis by Pyrosequencing. *Anal Biochem* 280:103–110

Chapter 4

Quantitative Validation and Quality Control of Pyrosequencing® Assays

Ulrich Lehmann

Abstract

Pyrosequencing® offers the opportunity to quantify the amount of nucleotides incorporated during the sequencing reaction. This enables the precise and sensitive quantification of DNA methylation, allele-specific expression, or mutation load. However, only a few studies addressed the question how precise and sensitive Pyrosequencing really is. Many publications claiming precise measurement (with two decimal places) fail to demonstrate performance figures like precision or reproducibility. And most publications reporting the detection (and quantification) of very weak signals by Pyrosequencing do omit to report the technical sensitivity of the assay used (i.e., limit of blank, limit of detection, limit of quantification).

Therefore, this chapter aims at providing (1) an introduction to important basic concepts defining quantitative analytical assays, (2) an outline of a minimal set of validation measurements every study employing Pyrosequencing should include, especially if values close to the analytical threshold are reported.

Key words Limit of detection, Limit of quantification, Quality control, Pyrosequencing®

1 Introduction

As many chapters in this volume and numerous other publications explain in great detail, Pyrosequencing enables among others the quantification of DNA methylation (*see* Chapters 13–22), allele-specific gene expression (*see* Chapter 12), or mutation load (*see* Chapters 5–8). However, laboratory assays producing a quantitative read-out require far more efforts for defining performance characteristics and validation steps than simple qualitative assays providing only a binary “yes” or “no” answer.

The following explanations intent to familiarize the Pyrosequencing user with basic analytic concepts characterizing qualitative and quantitative laboratory assays. The aim is to provide guidelines for estimating the quantitative reliability and capacity of the assays in use. It was not intended to provide a coherent mathematical deduction of all concepts and formulas. The reader interested in more details should be aware that key concepts (e.g., “limit of

detection”, LoD) are defined in slightly different versions in the literature and that even various regulatory agencies involved in the standardization of quantitative measurements in clinical laboratories define and calculate these parameters in slightly different ways (for more details *see* ref. [1] and references therein).

One reason for these inconsistencies is the often forgotten truth that the efforts one has to invest for defining the performance characteristics of a quantitative analytical assay depend substantially on what one wants to show and what one has to show. For example, the physiological and pathological levels of glucose or cholesterol in the blood in humans are so far above the detection threshold of the nowadays routinely applied analytical tools that it is not necessary to exactly determine the limit of detection for these clinically very important parameters. And if the mean methylation level of a gene of interest in patients is around 75 % and in controls is around 5 %, then the precision of the measurements has not to be determined with two decimal places. However, if the difference is only 2 % points or if the measured methylation levels are close to zero, then the precision of the assay and the limit of detection have to be determined.

The reader (and user) should also keep in mind that the performance of Pyrosequencing is heavily sequence context-dependent. That means achieving an analytical sensitivity of, for example, 3.5 % with a specific assay does not mean that Pyrosequencing in general has a sensitivity of 3.5 % (for the various and sometimes confused meanings of “sensitivity,” *see* **Note 1**). Another assay can do much worse depending on the sequence context. The CE-certified Pyrosequencing kit from Qiagen for the detection of clinically relevant EGFR mutations in non-small cell lung cancer (NSCLC) specimens, for example, reports in the accompanying handbook (TheraScreen® EGFR Pyro Kit Handbook) a LoD value of 10.7 % for the resistance-conferring point mutation EGFR T790M.

Also the precision of quantification is sequence-context dependent. If, for example, a potentially mutated site is part of a homopolymer stretch, precise quantification of alternatively incorporated nucleotides (indicating a mutation) might be difficult or even impossible (*see* Chapter 6 for an example).

2 Basic Concepts Defining Quantitative Analytical Tools

Every analytical assay intended to detect qualitatively the presence of an analyte or to measure the amount of an analyte in quantitative terms has a technically determined lower limit of detection below which measurements are not possible. Background signals (“noise”) in the absence of the analyte are an important parameter contributing to the existence of a lower limit for meaningful

measurements. Concerning Pyrosequencing accidental misincorporation of nucleotides and uncoupling of light signal generation from nucleotide incorporation due to hydrolysis effects are major sources of background signals preventing the reliable detection of very small amounts of the analyte (i.e., DNA methylation level, allelic load, proportion of allele-specific gene expression). Due to saturation effects most quantitative analytical assays also have an upper limit of measurement. This is, however, not an issue with Pyrosequencing, where the range of detection is by definition 0–1 or 0–100 %. Overloading of the Pyrosequencing reaction is prevented by the amount of Sepharose beads used that limits the number of template molecules transferred into the Pyrosequencing reaction.

From the relevant literature [1–6] one can extract the following definitions:

**2.1 Limit
of Blank (LoB)**

Highest value expected to see in a series of measurements on samples, which are negative for the signal of interest (e.g., completely unmethylated or pure wild-type sequence).

**2.2 Limit
of Detection (LoD)**

Lowest level of the signal of interest (e.g., mutated or methylated allele), which can be detected with sufficient confidence.

**2.3 Limit
of Quantification (LoQ)**

Lowest concentration of the signal of interest (e.g., mutated or methylated allele), which can reliably be measured in quantitative terms.

It will always be the case that $\text{LoB} < \text{LoD} \leq \text{LoQ}$.

Assuming that the blank values are normally distributed (*see Note 2*) and setting the 95th percentile as threshold, the LoB can be calculated as follows:

$$\text{LoB} = \text{mean}_{\text{Blank}} + 1.645 \times \text{sd}_{\text{Blank}}.$$

Under the assumption that the blank values are normally distributed and give a non-zero standard deviation a widely used approximation states:

$$\text{LoD} = \text{mean}_{\text{Blank}} + 3 \times \text{sd}_{\text{Blank}} \quad \text{and} \quad \text{LoQ} = \text{mean}_{\text{Blank}} + 10 \times \text{sd}_{\text{Blank}}.$$

See Note 3.

The calculations in Table 1 illustrate the consequences of these definitions and should supply the reader with a good deal of scepticism against claims that a study shows significant differences between patients and controls because the former display 2.5 % DNA methylation whereas the latter harbor only 1.4 % as measured by Pyrosequencing, because under many circumstances these values will be below the limit of quantitative detection and cannot be measured in any reliable way using this technology.

Table 1
Exemplary calculations of LoD and LoQ values for given blank values
(mean and standard deviation, SD)

Exemplary calculations for Limit of Detection (LoD) and Limit of Quantification (LoQ)			
Mean(blank)	SD(blank)	LoD	LoQ
0.5	0.2	1.1	2.5
0.5	0.4	1.9	4.5
1.0	0.3	1.9	4.0
1.0	0.6	2.8	7.0
2.0	0.3	2.9	5.0
2.0	0.6	3.8	8.0

These data clearly demonstrate that it will be very difficult (and under many circumstances simply not possible) to quantify methylation levels or allelic ratios below a level of 5 %

2.4 Linearity

Linear response between input of quantitatively defined calibration samples and analytical read-out.

2.5 Precision and Accuracy

Precision: how close are repeated measurements to each other?

This can be subdivided into within-run precision (scatter of intrarun replicates), run-to-run precision (scatter of repeatedly measured calibration samples), interlaboratory precision (scatter of values obtained from a homogeneous calibration sample in different laboratories using the same assay) (*see Note 4*).

Accuracy: how close are the measurements to the true value?

Due to amplification bias a very precise measurement can be quite inaccurate (far away from the true value), and due to large scattering of the data an accurate assessment can have a low precision (the mean is very close to the true value, but the standard deviation is quite large).

3 Practical Guidelines

As a minimum every study employing Pyrosequencing, it is recommended to incorporate the following controls:

A negative control (wild-type sequence or unmethylated sequence) has to be close to zero.

A positive control (fully mutated or methylated) has to be close to 100 %.

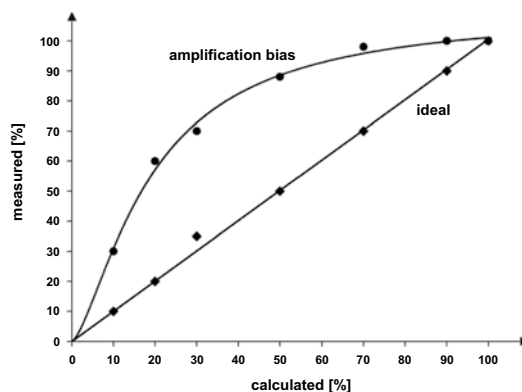


Fig. 1 By constructing a simple x/y -plot from the calculated and measured methylation values or allelic ratios derived from samples with defined composition, one can easily demonstrate the linearity of the assay or identify substantial amplification bias, which will severely distort the quantitative read out of the Pyrosequencing assay. As an example an amplification bias leading to overestimation of DNA methylation levels or allele frequencies is indicated. Sometimes the reverse is observed: methylated or mutated DNA is less efficiently amplified, leading to underestimation

Several samples with defined methylation level or allelic ratio should give a value close to the calculated one in order to ensure linearity of the assay.

A simple x/y -axes plot demonstrates the linearity (or the deviation from it) (*see* Fig. 1).

Especially for methylation analyses several studies have documented that sometimes either methylated or unmethylated alleles are amplified more efficiently (“amplification bias”) thereby distorting the quantitative read-out of the Pyrosequencing assays [7–9]. But the same phenomenon can be observed analyzing mutations on the DNA level or allelic ratios on the cDNA level.

Pyrosequencing offers the opportunity to control for the completeness of bisulfite treatment by introducing control dispensation at non-CpG cytosine positions, which should be completely converted to thymine. That means the dispensation of “C” should generate no signal at all. This is a clear advantage in comparison to many other not sequencing-based methods still widely used in the DNA methylation field, like methylation specific PCR (MSP), quantitative MSP (qMSP), or combined bisulfite restriction analysis (COBRA). Therefore, every Pyrosequencing assay should include several bisulfite treatment controls (*see* also Chapter 13 for further discussion of this point).

In case the obtained Pyrosequencing values are below 10 or even 5 % it is mandatory to determine LoB, LoD, and LoQ values

of the assays in use. The number of samples, which have to be analyzed for calculating these parameters varies in the literature. Armbruster and Pry [5], for example, recommend 60 replicates for establishment and 20 for verification. Under many circumstances 10–20 measurements might be sufficient. In order to ensure that the control values are truly representative, one should combine measurements from several different control samples and not measure a single control 10 or 20 times.

4 Sample Types for Controls

Table 2 gives an overview about the various sources of positive and negative controls.

For DNA methylation studies one can also purchase completely methylated or unmethylated human genomic DNA (for potential sources of supply, *see Note 5*). Alternatively, one can prepare these DNAs themselves using *SssI* methylase for complete methylation and Phi29 polymerase for the generation of completely unmethylated DNA. However, in our hands both commercially available fully methylated DNA and in house *SssI*-methylated DNA is never 100 % methylated in all loci looked at.

In our hands, DNA isolated from fresh-frozen samples and DNA isolated from FFPE samples behave sometimes not identical. Therefore, the calibration of the assay has to be performed with the sample type, which will be analyzed in the research project or

Table 2
Sources of controls

	Advantages	Disadvantages
Patient sample	Resembles real samples best	Often not available or only in small limited amounts Tissue samples are heterogeneous Composition not standardized
Plasmid	100 % pure and defined Total flexibility Unlimited supply	One million times less complex than human genome Amplification of plasmid does not match amplification of complex DNA samples
Cell line	Very similar to real samples Unlimited supply	Often not available Not as homogeneous as assumed
In vitro methylated DNA	Commercially available Unlimited supply	Not 100 % homogeneously methylated Commercial version quite expensive In house version requires hands on-time

The table gives an overview about all sorts of positive and negative controls and the respective advantages and disadvantages

in routine diagnostics. Differentially processed patient samples are not always directly comparable.

Under many circumstances it is not necessary to determine all above-outlined performance figures for every Pyrosequencing assay in use. However, if one claims that the mean methylation level of a gene in patients is 3.1 % and statistically significantly different from a mean value of 1.9 % in controls, one has to determine the limit of quantitative detection and the precision of the assay in use. Furthermore, a Pyrosequencing assay used for diagnostic purposes should have a very similar performance in different laboratories.

A totally different quality control issue well beyond the topic of this contribution is the cellular composition of the sample under study. Very interesting recent publications (*see* ref. [10] and references therein) clearly demonstrate that many quantitative differences in DNA methylation reported in the literature are due to changes in cellular composition and not due to genuine changes in DNA methylation level.

5 Notes

1. Under many circumstances “sensitivity” refers to the proportion of positive results correctly identified by a laboratory test. Accordingly, a value close to 100 % is the desirable. This term must not mixed up with “analytical sensitivity” referring to the minimal concentration, amount, or proportion that can be detected or measured reliably. In this situation a value as small as possible is desirable.
2. Even if the measured values are not normally distributed by convention this is very often assumed, because it facilitates calculations and the resulting error is under many circumstances negligible.
3. For those readers familiar with real-time PCR the formula $\text{mean}_{\text{Background}} + 10 \times \text{sd}_{\text{background}}$ during cycle no. 3–15 has been used by ABI (now ThermoFisherScientific) since the Mid-1990s for the calculation of the threshold of detection for C_T -value determination in real-time PCR measurements.
4. The precision depends also on the fluctuation of the baseline. According to our experience the baseline of the PyroMark MD system fluctuates by 3–4 relative light units (RLU), largely independent from the signal strength for single nucleotide incorporation and across many different assays and different assay types. This can translate into a signal variance due to baseline fluctuations of several percentage points!
5. Examples for commercial sources of fully methylated and fully unmethylated DNA: Active Motif, Zymo Research, New England Biolabs, or ThermoFisherScientific/Life Technologies.

Acknowledgement

The author would like to thank Prof. Dr. Harald Binder, IMBEI, Mainz, and Dr. Norbert Hochstein from Qiagen for interesting and insightful discussions about several topics presented in this chapter.

References

1. Shrivastava A, Gupta VB (2011) Methods for the determination of limit of detection and limit of quantification of the analytical methods. *Cytosonline* 2:21–25
2. Currie LA (1968) Limits for qualitative detection and quantitative determination. *Anal Chem* 40:586–593
3. Haeckel R, Baud M (1994) Detection limit, a linguistic problem? *Eur J Clin Chem Clin Biochem* 32:43–44
4. Lindstrom RM (2001) Limits for qualitative detection and quantitative determination. *Nat Inst Sci Technol, NIST SP-958*, 164–166
5. Armbruster DA, Pry T (2008) Limit of blank, limit of detection and limit of quantitation. *Clin Biochem Rev* 29(Suppl 1):S49–S52
6. Proctor CH (2008) A simple definition of detection limit. *J Agric Biol Env Stat* 13:1–23
7. Warnecke PM, Stirzaker C, Melki JR et al (1997) Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. *Nucleic Acids Res* 25:4422–4426
8. Tost J, Dunker J, Gut IG (2003) Analysis and quantification of multiple methylation variable positions in CpG islands by Pyrosequencing. *Biotechniques* 35:152–156
9. Moskalev EA, Zavgorodnij MG, Majorova SP et al (2011) Correction of PCR-bias in quantitative DNA methylation studies by means of cubic polynomial regression. *Nucleic Acids Res* 39:e77
10. Jaffe AE, Irizarry RA (2014) Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol* 15:R31

Part II

Analysis of Genetic Variation

Free ebooks ==> www.Ebook777.com

Chapter 5

Extended *KRAS* and *NRAS* Mutation Profiling by Pyrosequencing®

Andreas Jung

Abstract

With the advent of targeted therapies—drugs that specifically target molecules of tumor-driving signalling pathways—and the availability of biomarkers that predict the response of an individual patient on such a targeted therapy, the analysis of the status of the biomarker became an integral part of the therapy. For metastatic colorectal cancer, anti-EGFR-targeted antibodies (Cetuximab/Erbitux®, Panitumumab/Vectibix®) fall into this category of drugs as it was shown in several clinical studies that oncogenic mutations in exons 2–4 of the RAS genes *KRAS* or *NRAS* result in therapeutic resistance of the metastatic colorectal cancers against the action of both targeted drugs. Therefore, mutations in the RAS genes exclude patients from this kind of targeted therapy (negative biomarker). Thus the molecular-pathological testing of the mutational status of *KRAS* and *NRAS* has become an important cornerstone in planning oncological strategies in the treatment of metastatic colorectal cancer. As the profile of mutations in the RAS genes is characterized by hotspot mutations in only a small number of codons (12, 13 in exon 2–59, 61 in exon 3–117, 146 in exon 4). Pyrosequencing® is an ideal and robust tool in the molecular-pathological detection. A detailed protocol for this detection procedure employing Pyrosequencing® is given here.

Key words Pyrosequencing®, Colorectal cancer, anti-EGFR-targeted therapy, *KRAS*, *NRAS*, RAS-hypothesis

1 Introduction

Since about 2007 predictive and individualized medicine based on the molecular-pathological analysis of human tissues has become an integral part of the treatment of tumour patients with colorectal cancer (CRC) [1]. The conception of predictive and individualized therapy is based on the diagnostics (also known as theranostics [2]) of predictive biomarkers [3]. For the finding of both targets as well as biomarkers the knowledge and understanding of signalling pathways on the molecular level plays a central role. Obviously, targetable molecules in those signalling pathways that have an important role in driving the process of carcinogenesis are the focus of interest [4, 5]. In CRC the epidermal growth factor receptor (EGFR)—rat sarcoma (RAS)/RAS-associated factor (RAF)/mitogen-activated

protein kinase (MAPK) signalling pathway is an essential driver [6]. EGFR is a member of the receptor tyrosine kinase (RTK) family. RTKs are activated by ligand-mediated receptor dimerization. In CRC the relevant ligands for the EGFR are epiregulin (ER) and amphiregulin (AR). Thus, antibodies like cetuximab (C-Mab, Erbitux®) or panitumumab (P-Mab, Vectivix®), which are able to out-compete the binding of the ligands AR and ER to EGFR, are effective molecular drugs, also known as biologicals, and able to shut down the activity of the EGFR/RAS/RAF/MAPK signalling pathway. According to the RAS-hypothesis [1], this works only if there is no downstream activation in this pathway at the same time as such activation will result in an autonomous activation of the pathway downstream of the point of inhibition at the receptor level. Such activations occur in about 50 % of all cases of CRCs through mutations of the two RAS family members Kirstin-RAS (*KRAS*) and neuronal RAS (*NRAS*) by hotspot mutations mostly in their codons 12 and 13 (exon 2), 59 and 61 (exon 3), and 117 and 146 (exon 4) (Fig. 1).

The RAS-hypothesis was verified in several clinical studies in the first step for *KRAS* exon 2 [1, 7–9] and later also for the RAS genes (*KRAS* and *NRAS*) [10]. Thus, only patients with wild-type (WT) RAS-genes and thus proteins benefit from the targeted anti-EGFR therapy whereas patient with mutant (MUT) RAS either do not profit from the therapy or are even harmed. Therefore, the European Medicines Agency (EMA) approved anti-EGFR-directed therapies with monoclonal antibodies only for patients without mutant RAS-genes (<http://www.ema.europa.eu/ema/>). In addition, when comparing the effects of anti-EGFR-directed C-Mab or P-Mab with antibodies directed against vascular endothelial growth factor (anti-VEGF, bevacizumab—B-Mab, Avastin®) in the clinical studies PEAK [11] and FIRE3 [12], it turned out that the action of the anti-EGFR-targeted biologicals out-performed the anti-VEGF-targeting B-Mab with respect to overall survival (OS) but only in patients with WT RAS-loci. Thus the German working group intestinal oncology (AIO) recommended the combination of chemotherapy with anti-EGFR-targeting antibodies for the first-line treatment of patients with CRC (<http://www.aio-portal.de/index.php/stellungnahmen.html>).

Thus, one important task in the treatment of CRC is nowadays the molecular-pathological determination of the mutational status of the RAS-loci in the tumour cells. As the mutations in the RAS-genes are typically found in hotspots located next to each other (see for example the COSMIC database, <http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/>), they represent an ideal target for the application of the Pyrosequencing® technique. Protocols for the detection of mutations in the exons 2–4 of both oncogenes *KRAS* and *NRAS*, which have been proven to work successfully in daily practice are presented in this chapter.

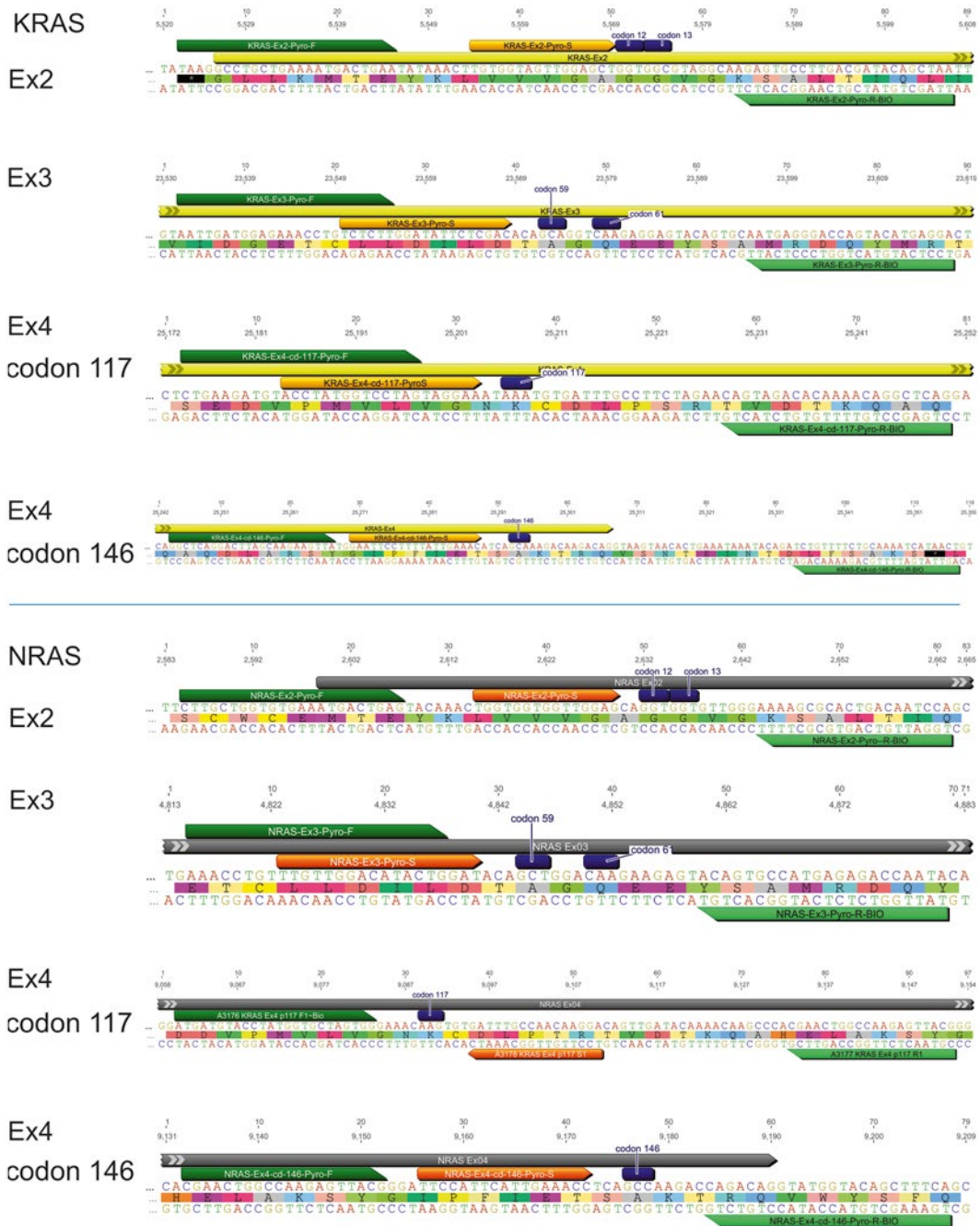


Fig. 1 Localization of the primers used for Pyrosequencing of the RAS genes. The positions of the genes are numbered according to the genomic sequences of *KRAS* (NC_000012.11—position: 25403854-25358180) and *NRAS* (NC_000001.10—position: 115261379-115245219). Forward primers are depicted in dark green, reverse primers in light green, sequencing primers in orange, codons in blue, and exons in yellow (*KRAS*) or grey (*NRAS*)

2 Materials

2.1 Assay Design

1. PSQ assay design software (Qiagen).

Usually, two primer systems are tested in parallel. Besides the well-known thermodynamic criteria another important main criterion—when working with formalin fixed paraffin embedded (FFPE) tissue—is the selection of primer pairs that result in short-sized amplicons, best between 50 and 100 bp length.

2.2 Sample Preparation, Calibration Standards

1. Pyro-Control Oligo (Qiagen).
2. QIAamp® DNA FFPE kit, QIAamp DNA Mini kit (Qiagen).
3. QiaCube device (Qiagen).
4. Male DNA (Promega).

2.3 PCR Amplification

1. dNTPs—store at 4 °C.
2. HotStar Taq-DNA Polymerase or similar—store at –20 °C.
3. Mastercycler® Pro thermocycler or similar.
4. Primer-pairs for PCR-amplification with one of the primers carrying a biotin-group at its 5' end—store at 4 °C.

2.4 Sample Preparation for Pyrosequencing Analysis

1. 24-PCR plate (Brand, Wertheim, Germany).
2. Annealing Buffer (Qiagen)—store at 4 °C.
3. Binding Buffer (Qiagen)—store at 4 °C.
4. Denaturation Solution (Qiagen)—store at 4 °C.
5. Heating plate.
6. HPLC grade water.
7. Monoshake Mixer.
8. Pyro Troughs (Qiagen).
9. Pyro Vacuum Prep Station (Qiagen).
10. PyroMark® Q24 Plate 100, version 1 (Qiagen).
11. Sequencing Primers—store at 4 °C.
12. Streptavidin Sepharose HP beads (GE Healthcare)—store at 4 °C.
13. Washing Buffer (Qiagen)—store at 4 °C.

2.5 Analysis by Pyrosequencing Reaction

1. PyroMark Gold Q24 (Qiagen).
2. PyroMark Q24 Cartridge (Qiagen).

3 Methods

3.1 Assay Design

Any software can be used to select for appropriate primers. It is recommended to use software tools, which are based on the algorithm of Rychlik [13, 14], which takes several thermodynamic

considerations into account resulting in less primer dimers or other by-products in the PCR.

- Low ΔG (free Gibbs energy) in the middle part (CG rich sequence).
 - Higher ΔG at the 3' end of the primer (AT rich sequence).
 - 3' base is a C or G.
 - No palindromic sequences avoiding hairpin structures.
 - No alternative binding sites in the genome avoiding unwanted and unspecific binding.
 - PCR products should be in the range of 50–100 bp when DNA from FFPE-tissue is used due to the chemical modification (Schiff's base) of DNA by formalin.
1. Use NCBI's open-source search tool Gene (<http://www.ncbi.nlm.nih.gov/gene/>) for identifying sequence files:

<i>KRAS</i>	
g-DNA:	NC_000012.11—position: 25403854-25358180
mRNA/cDNA:	NM_004985
<i>NRAS</i> :	
g-DNA:	NC_000001.10—position: 115261379-115245219
mRNA/cDNA:	NM_002524

2. Use NCBI's open-source alignment tool Spidey (<http://www.ncbi.nlm.nih.gov/spidey/>) for finding exon/intron boundaries in the g-DNA.
3. Localize the region containing the mutation in the cDNA by calculating the codons (C) namely the first (N_1^C) and the last base (N_3^C) of the codons of the involved region applying the formulas:

$$N_1^C = 3 \times C - 2 \quad (1)$$

$$N_3^C = 3 \times C \quad (2)$$

Thus, the region spanning codons 12 and 13 were calculated as $N_1^{12} = 3 \times 12 - 2 = 34$ and $N_3^{13} = 3 \times 13 = 39$ resulting in the region of bases 34–39.

4. Transfer this region onto the genomic DNA sequence.
5. Select 100 bp both upstream and downstream of the defined region of interest and copy this DNA stretch into the PSQ assay design software.
6. Select the region of interest (CTRL T).
7. Run the program and select appropriate primers.

Table 1
Primer sequences

Assay	Primer name	Sequence
<i>KRAS</i> exon 2	KRAS-Ex2-Pyro-F	NNNGGCCTGCTGAAAATGACTGAA
	KRAS-Ex2-Pyro-R-BIO	BIO~TTAGCTGTATCGTCAAGGCACTCT
	KRAS-Ex2-Pyro-S	TGTGGTAGTTGGAGCT
<i>KRAS</i> exon 3	KRAS-Ex3-Pyro-F	AATTGATGGAGAAACCTGTCTCTT
	KRAS-Ex3-Pyro-R-BIO	BIO~TCCTCATGTACTGGTCCCTCATT
	KRAS-Ex3-Pyro-S	TCTCTTGGATATTCTCGAC
<i>KRAS</i> exon 4 codon 117	KRAS-Ex4-cd-117-Pyro-F	CTGAAGATGTACCTATGGTCCTAG
	KRAS-Ex4-cd-117-Pyro-R-BIO	BIO~CTGAGCCTGTTTTGTGTCTACTG
	KRAS-Ex4-cd-117-Pyro-S	ACCTATGGTCCTAGTAGGAA
<i>KRAS</i> exon 4 codon 146	KRAS-Ex4-cd-146-Pyro-F	GGCTCAGGACTTAGCAAGAAGTTA
	KRAS-Ex4-cd-146-Pyro-R-BIO	BIO~AGTTATGATTTTGCAGAAAACAGA
	KRAS-Ex4-cd-146-Pyro-S	GAATTCCTTTTATTGAAAC
<i>NRAS</i> exon 2	NRAS-Ex2-Pyro-F	CTTGCTGGTGTGAAATGACTGAG
	NRAS-Ex2-Pyro-R-BIO	BIO~TGGATTGTCAGTGCCTTTT
	NRAS-Ex2-Pyro-S	TGGTGGTGGTTGGAG
<i>NRAS</i> exon 3	NRAS-Ex3-Pyro-F	AAACCTGTTTGTGGACATACTG
	NRAS-Ex3-Pyro-R-BIO	BIO~TATTGGTCTCTCATGGCACTGT
	NRAS-Ex3-Pyro-S	TTGTTGGACATACTGGAT
<i>NRAS</i> exon 4 codon 117	NRAS-Ex4-cd-117-Pyro-F-BIO	BIO~ATGATGTACCTATGGTGCTAGTGG
	NRAS-Ex4-cd-117-Pyro-R	CGTAACTCTTGGCCAGTTTCG
	NRAS-Ex4-cd-117-Pyro-S	TCCTTGTTGGCAAATC
<i>NRAS</i> exon 4 codon 146	NRAS-Ex4-cd-146-Pyro-F	CGAACTGGCCAAGAGTTACG
	NRAS-Ex4-cd-146-Pyro-R-BIO	BIO~TGAAAGCTGTACCATACCTGTCTG
	NRAS-Ex4-cd-146-Pyro-S	TCCATTCATTGAAACCT

Primer sequences for the eight loci covering the mutation hot spot for the occurrence of mutations in the *KRAS* and *NRAS* oncogenes. Oncogenic mutations in the RAS genes are found in codons 12 and 13 (exon 2), 59 and 61 (exon 3), and 117 and 146 (exon 4) of *KRAS* as well as *NRAS*

8. Use NCBI's open-source alignment tools BLASTN (<http://blast.ncbi.nlm.nih.gov/>) to check for alternative binding sites of the primers in the genome like pseudogenes. Only specific primers are used (Table 1) (*see* Note 1).
9. After optimization (*see* Note 2) eight primer pairs were selected for the detection of mutations in the exons 2–4 for the *KRAS*- and *NRAS* gene in metastatic colorectal cancer (Fig. 2) (*see* Note 3).
10. Write down the sequence to analyse (Table 2) considering all relevant mutations (COSMIC).
11. Generate dispensations orders (Table 2) using the PyroMark Q24 software.

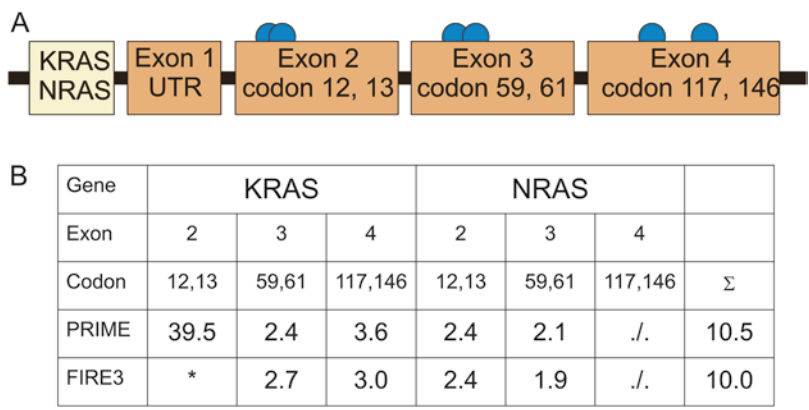


Fig. 2 Organization of the *KRAS* and *NRAS* gene loci and distribution of mutations. (a) The RAS genes *KRAS* and *NRAS* are homologues and are therefore more or less identically organized. The coding region (CDS) is found in the three exons 2, 3, and 4. Mutations are found in the codons 12, 13 (exon 2); 59, 61 (exon 3); and 117, 146 (exon 4). (b) The frequencies of the mutations (given in %) were found in the Prime [10] and FIRE3 [12] studies. About 40 % of the patients with metastatic colorectal cancer harbour mutations in *KRAS* exon 2 (COSMIC) and additional 10 % in the two RAS genes other than *KRAS* exon 2. Taken together about 50 % of patients with this disease display mutations in one of the three exons of the RAS genes

Table 2
Dispensation order

Assay	Sequence to analyse	Dispensation order
<i>KRAS</i> exon 2	GNTGRCGTAGG	CATGGACTGGACGTAG
<i>KRAS</i> exon 3	AYAKCAGGWCAHGAGGAGTA	GACAGCAGGTACGATCAGAGGAGTA
<i>KRAS</i> exon 4 codon 117	ATAAHTGTG	CAGTACTGTG
<i>KRAS</i> exon 4 codon 146	ATCAVYAAAGA	CATGCAGCTAGA
<i>NRAS</i> exon 2	CAYGTYGTGTT	TCGATGACTCGATGT
<i>NRAS</i> exon 3	ACAGCTGGACDASAGAG	CACAGCTGAGCGTACGAG
<i>NRAS</i> exon 4 codon 117	ACACNTGTTTCCC	CACGACAGTGTC
<i>NRAS</i> exon 4 codon 146	TCARCCAAG	CTGCAGCAG

Based on the sequence to analyse dispensations orders were generated employing the PyroMark Q24 software. Ambiguities are encoded by the IUPAC (International Union of Pure and Applied Chemistry) system: **B** (C/G/T), **D** (A/G/T), **H** (A/C/T), **K** (G/T), **M** (A/C), **N** (A/C/G/T), **R** (A/G), **S** (C/G), **F** (A/C/G), **W** (A/T), **Y** (C/T)

**3.2 Sample
Preparation,
Calibration Standards**

The biological material for the RAS-gene analyses is FFPE tissue from patients with metastatic colorectal cancer. Thus, it is essential to do a histomorphological inspection of the tissue ahead of any analysis to warrant that the amount of tumour cells is sufficient for the mutation test.

1. Prepare several serial sections (2–3 μm) of the FFPE tumour tissue block of interest.
2. Stain one of the sections using H&E.
3. A histo-pathologist marks areas containing sufficient amount of tumour cells (>25 %, at least 200 cells).
4. Dewax the remaining sections.
5. Transfer the marked area onto the unstained section.
6. Scratch off the tissue from this area by macroscopic micro-dissection—that is scratch off the tissue with the help of a scalpel blade.
7. Transfer the tissue into a micro-reaction tube containing 200 μL lysis buffer (180 μL ATL buffer, 20 μL proteinase K).
8. Incubate the tissue in the lysis buffer at 56 °C under permanent shaking between 30 min till overnight.
9. Isolate the DNA from the lysate:
 - (a) Manually: follow the steps given in the handbooks.
 - Step 12—QIAamp DNA FFPE kit Tissue Handbook 04/2010.
 - Step 3—QIAamp DNA Mini kit tissue protocol.
 - (b) Automatically: following the protocol of the QiaCube.
10. Elute the DNA in a volume of 20 μL elution Buffer (EB) (*see Note 4*).

3.3 PCR Amplification

1. Pipet the master mix best in a clean hood (*master mix hood*). Consider also two controls (negative—and positive control) (*see Note 5*).

	Concentration		Master mix	
	Stock	Final	1 $\times$ [μL] ²	x $\times$ [μL] ³
	Stock	Final	1 $\times$ [μL] ²	x $\times$ [μL] ³
H ₂ O			32.2	...
10 $\times$ Hot Star PCR buffer ¹	10 $\times$	1 $\times$	4.0	...
dNTP	10 mM	0.2 mM	0.8	...
Primer-Mix	20 μM	0.4 μM	0.8	...
Qiagen Hot Start Taq-Polymerase	5 U/ μL	1 U	0.2	...
			38.0	
DNA			2.0	./.

2. Mix the master mix, spin down briefly (micro-centrifuge).
3. Transfer 38 μL of master mix in PCR microreaction tubes.
4. Add 2 μL of water to the first PCR microreaction tube and close it.
This negative control serves as a control of the master mix and the *master mix hood*.
5. Transfer the prepipetted PCR microreaction tubes in the second hood (DNA hood) and add 2 μL of template DNA of the prepared samples to the master mix and close the tube immediately.
6. Pipet in the second but last PCR micro-reaction tube 2 μL of water and close the tube.
This negative control serves as a control of the *DNA hood* and spill over from pipetting.
7. Finally, pipet 2 μL of male DNA (10 ng/ μL) into the last PCR micro-reaction tube.
This positive control controls for testing the functionality of the PCR.
8. Mix, spin down briefly (microcentrifuge).
9. Run PCR applying the PCR profile (*see Note 6*):

Cycle type	Cycle no.	Cycle profile
Activation of Taq-polymerase	1×	15 min—95 °C
PCR cycles	50×	30s—95 °C, 30s—60 °C, 30s—72 °C
Final elongation	1×	2 min—72 °C, 1 s—20 °C—off

3.4 Sample Preparation for Pyrosequencing Analysis

A comprehensive description is given in the PyroMark Q24 User Manual 01/2009.

1. Prepare the run using PyroMark Q24 software.
Alternative to a manual input the names (ID) and additional information (add info) of the probes can be imported from a CSV-file with the following structure:

Well	ID	Add info
A1		
A2		
...		
C8		

2. Save the new run and transfer it to an USB stick.
3. Calculate the amount of reagents needed for the reaction (Tools → Pre Run Information).
4. Calculate the amount of streptavidin-coated beads in binding buffer as follows (*see* **Note 7**):

	Master mix	
	$1 \times^2$	x^3
Beads ¹	2	...
Binding buffer	40	...
PCR product ⁴	15	./.
Water	23	...
	80	

5. Calculate the amount of sequencing primers in annealing buffer.

First, prepare a $10 \times$ ($3 \mu\text{M}$) sequencing primer solution in annealing buffer.

Second, prepare a master mix of the sequencing primers as follows (*see* **Note 8**):

	Concentration		Master mix	
	Stock	Final	$1 \times [\mu\text{L}]^1$	$x \times [\mu\text{L}]^2$
Sequencing primer	$3 \mu\text{M}$	300 nM	2.5	...
Annealing buffer			22.5	...
			25.0	

6. Thaw the PyroMark Gold Q24 reagents (enzyme and substrate).
7. Fill the troughs with sufficient amount of buffers as given in the handbook of the VacuPrep station.
8. Prepare the master mixes for the beads and the primers.
9. Transfer $65 \mu\text{L}$ of the beads master mix into a V-bottom plate according to the run protocol. Add $15 \mu\text{L}$ of the respective PCR product.
10. Transfer the V-bottom plate in the shaker and mix vigorously for 3 min (position 3 o'clock of the turning knob).
11. Transfer $25 \mu\text{L}$ of the respective sequencing primer into a flat bottom plate according to the run protocol. And transfer this plate into its position in the VacuPrep station.

12. Remove the V-bottom plate from the shaker and transfer it to its position in the VacuPrep station.
13. Switch on the pump to establish low-pressure in the evacuation system. Test the low-pressure by holding the comb into one of the water containers. If water appears in the flexible tube the system is ready.
14. Put the comb into the V-bottom plate position and soak off the fluid containing the beads for 15 s.
15. Transfer the comb subsequently into:
 - 70 % Ethanol for 5 s.
 - Denaturation solution for 5 s.
 - Washing buffer for 10 s.
16. Hold the comb in a vertical position and move it waving for 5 s.
17. Carefully transfer the comb in the position of the flat-bottom well containing the primers and turn-over the rocker switch. Wait some seconds and then lower the comb into the primer mix. Shake the comb to release the beads into the solution containing the primers (*see Note 9*).
18. Wash the comb in water for 10 s and transfer it into the next water-containing trough for cleaning the unit. Therefore, close the valve by turning the rocker-switch in the ON position.
19. Transfer the flat-bottom well plate onto a PyroMark Q24 plate holder and incubate for 2 min at 80 °C.
20. Fill the PyroMark Q24 cartridge according to the prerun information.
21. Remove the flat bottom plate from the heater and let it cool down for about 30 s.
22. Transfer the loaded PyroMark Q24 cartridge into the PyroMark Q24 device.
23. Transfer the flat bottom plate into the PyroMark Q24 device and close the lid.
24. Run the program.
25. Analyze the data employing the PyroMark Q24 software.

4 Notes

1. Usually it is recommendable to try two independent primer systems in the beginning and select then the better one for routine.
2. Optimization includes finding of:
 - (a) The best melting or annealing temperature (T_M) of the primer-pairs,

- (b) The specificity of the primer systems by increasing the cycle numbers in ten steps from 40 to 60 [13],
 - (c) The sensitivity of the individual primer-pairs using dilutions of DNAs from cell lines carrying mutations in the region of investigation and dilutions of FFPE tissue.
3. As the regions for codons 117 and 146 of both the KRAS and NRAS genes were too much separated from each other two independent pyro-sequencing assays were generated.
 4. As the DNA is chemically modified by formaldehyde (Schiff's bases) and no essential hybridization steps are part of the protocol the DNA concentration was not measured as it is a more or less useless factor.
 5. ¹Qiagen's 10× PCR buffer already contains 15 mM MgCl₂ resulting in a final concentration of 1.5 mM MgCl₂.
²The volumes in the column 1× represent the amount of reagent needed for a single reaction.
³Multiply the digits in the column 1× with the number of reactions (*x*). This will result in the total volumes for the master mix.
 6. *PCRs can be run at least up to 60 cycles without introducing errors [15].
 7. ¹Mix the beads thoroughly before use to produce homogeneous slurry.
²The volumes in the column 1× represent the amount of reagent needed for a single reaction.
³Multiply the digits in the column 1× with the number of reactions (*x*). This will result in the total volumes for the master mix.
⁴The PCR product is not part of the master mix. It is just given for reasons of calculation.
 8. ¹The volumes in the column 1× represent the amount of reagent needed for a single reaction.
²Multiply the digits in the column 1× with the number of reactions (*x*). This will result in the total volumes for the master mix.
 9. If the comb is moved to quickly into the flat-bottom plate containing the primer mix it can happen that the primer mix is soaked off accidentally as the reduction of the built-up vacuum requires a moment.

References

1. Amado RG, Wolf M, Peeters M et al (2008) Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. *J Clin Oncol* 26:1626–1634
2. Warner S (2004) Diagnostics+therapy=theranostics. *The Scientist* 18:38
3. Biomarker Working Group (2001) Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 69:89–95
4. Koopman M, Venderbosch S, Nagtegaal ID et al (2009) A review on the use of molecular

markers of cytotoxic therapy for colorectal cancer, what have we learned? *Eur J Cancer* 45:1935–1949

5. Winder T, Lenz HJ (2010) Molecular predictive and prognostic markers in colon cancer. *Cancer Treat Rev* 36:550–556
6. Schubbert S, Shannon K, Bollag G (2007) Hyperactive Ras in developmental disorders and cancer. *Nat Rev Cancer* 7:295–308
7. Bokemeyer C, Bondarenko I, Makhson A et al (2009) Fluorouracil, leucovorin, and oxaliplatin with and without cetuximab in the first-line treatment of metastatic colorectal cancer. *J Clin Oncol* 27:663–671
8. Douillard JY, Siena S, Cassidy J et al (2010) Randomized, phase III trial of panitumumab with infusional fluorouracil, leucovorin, and oxaliplatin (FOLFOX4) versus FOLFOX4 alone as first-line treatment in patients with previously untreated metastatic colorectal cancer: the PRIME study. *J Clin Oncol* 28:4697–4705
9. Van Cutsem E, Kohne CH, Hitre E et al (2009) Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer. *N Engl J Med* 360:1408–1417
10. Douillard J-Y, Oliner KS, Siena S et al (2013) Panitumumab–FOLFOX4 treatment and RAS mutations in colorectal cancer. *N Engl J Med* 369:1023
11. Schwartzberg LS, Rivera F, Karthaus M et al (2014) PEAK: a randomized, multicenter phase II study of panitumumab plus modified fluorouracil, leucovorin, and oxaliplatin (mFOLFOX6) or bevacizumab plus mFOLFOX6 in patients with previously untreated, unresectable, wild-type KRAS exon 2 metastatic colorectal cancer. *J Clin Oncol* 32:2240–2247
12. Heinemann V, Fischer von Weikersthal L, Decker T et al (2013) Randomized comparison of FOLFIRI plus cetuximab versus FOLFIRI plus bevacizumab as first-line treatment of KRAS wild-type metastatic colorectal cancer: German AIO study KRK-0306 (FIRE-3). *J Clin Oncol* 31, LBA3506
13. Rychlik W, Rhoads RE (1989) A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in vitro amplification of DNA. *Nucleic Acids Res* 17:8543–8551
14. Rychlik W, Spencer WJ, Rhoads RE (1990) Optimization of the annealing temperature for DNA amplification in vitro. *Nucleic Acids Res* 18:6409–6412
15. Vogelstein B, Kinzler KW (1999) Digital PCR. *Proc Natl Acad Sci U S A* 96:9236–9241

Free ebooks ==> www.Ebook777.com

Chapter 6

Universal BRAF State Detection by the Pyrosequencing®-Based U-BRAF^{V600} Assay

Alexander Skorokhod

Abstract

Malignant melanoma is a highly aggressive type of malignancy with considerable metastatic potential and frequent resistance to cytotoxic agents. BRAF mutant protein was recently recognized as therapeutic target in metastatic melanoma.

We present the newly developed U-BRAF^{V600} approach (Patent No. 12153477.0-1404)—a universal Pyrosequencing®-based assay for mutation detection within the activation segment in exon 15 of human *BRAF*. We identified five different *BRAF* mutations in a single assay analyzing 75 different formalin-fixed paraffin-embedded samples of cutaneous melanoma metastases from 29 patients. All mutant variants were quantitatively detectable by the newly developed U-BRAF^{V600} assay. These results were confirmed by ultra-deep sequencing. In contrast to all other *BRAF* state detection methods, the U-BRAF^{V600} assay is capable of automated quantitative identification of at least 36 previously published *BRAF* mutations. Under the precaution of a minimum of 5 % mutated cells in a background of wild-type cells, the U-BRAF^{V600} assay design completely excludes false-negative wild-type results. The corresponding algorithm for classification of BRAF-mutated variants is provided in this chapter together with a detailed step-by-step protocol for the Pyrosequencing reaction.

The single-reaction format and automation of data analysis make our approach suitable for the assessment of large clinical cohorts. Therefore, we suggest U-BRAF^{V600} assay as a powerful sequencing-based diagnostic tool to automatically identify *BRAF* status.

Key words FFPE tissue DNA, Malignant/metastatic melanoma, Hotspot BRAF V600E mutation, Pyrosequencing®

1 Introduction

Malignant melanoma is a highly aggressive skin cancer and one of the most metastatic malignancies [1]. Studies revealed that regulation of the Ras-Raf-MAPK pathway is abrogated in the majority of melanoma tumors as a result of activating *NRAS* or *BRAF* mutations, which are mutually exclusive and present in up to 90 % of cutaneous melanomas [2]. *BRAF* mutation is an early event in tumorigenesis: it may already occur in benign nevus, but by itself is not sufficient to induce cancer [3]. It was suggested that effects of

NRAS and *BRAF* mutations may be limited to early disease stages and that other factors are more relevant after regional metastases have occurred [4]. Somatic mutations in the *BRAF* oncogene have been documented with high frequency in cutaneous melanomas, occurring in 50–70 % of tumor samples [5]. *BRAF* mutations are also found in 40–70 % of papillary or anaplastic thyroid cancers and in small percentages of many other types of tumor [6]. Most *BRAF* mutations occur at codon V600 and constitutively activate BRAF together with the corresponding downstream signal transduction in the MAP kinase pathway [6]. This mutation significantly increases the risk of mortality both in colorectal cancer patients and in patients with malignant melanoma [7].

The BRAF kinase inhibitor vemurafenib, recently approved by the Food and Drug Administration (FDA), represents significant progress in melanoma therapy: patients' treatment with vemurafenib resulted in complete or partial tumor regression in the majority of patients with BRAF^{V600E}-positive metastatic melanoma [8].

Sanger (direct) sequencing is widely accepted as a gold standard routinely used to detect down to 20 % *BRAF* mutation level in biopsy specimens [9]. Alternative approaches, like the cobas® BRAF V600 Mutation Test (Roche) or BRAF RGQ PCR (Qiagen), claim to detect mutations down to 1.27 % level in a wild-type background. Nevertheless, as quantitative PCR-based approaches, they have limited precision and present difficulties in reliably detecting low-copy-number templates due to nonspecific amplification and competitive side reactions [10]. Unfortunately, the FDA-approved cobas 4800 BRAF V600 Mutation Test is not able to distinguish between mutations V600E, V600K, and V600E2 (*see Note 1*). Moreover, according to the FDA's Summary of Safety and Effectiveness Data (SSED), mutation levels of less than 30 % for V600K and 68 % for V600E2 mutations (c.TG1799_1800AA) are not detectable by the cobas BRAF V600 Mutation Test assay. *BRAF* mutation assays based on restriction fragment length polymorphism analysis (RFLP) and single-strand conformation polymorphism analysis (SSCP) are less sensitive and less specific than Sanger sequencing [11].

In contrast, Pyrosequencing®, a real-time sequencing-by-synthesis approach, has a high throughput and is capable of detecting minor sequencing variants with greater diagnostic sensitivity than Sanger sequencing (Table 1). It shows high accuracy and precision of Pyrosequencing for the quantitative identification of BRAF mutations in melanoma cell lines as well as in formalin-fixed paraffin-embedded (FFPE) tumors [12]. Even though the approaches based on shifted termination assay (STA) and amplification refractory mutations system allele-specific PCR (ARMS AS-PCR) give comparably sensitive results, they are designed for the detection of only very few *BRAF* mutation variants. In general, to avoid false-negative wild-type results, Sanger sequencing is required for

Table 1
BRAF mutations within activation segment in exon 15 in cutaneous melanoma metastases

Case	Sample	Age/sex	Sanger sequencing	Pyrosequencing mt:wt ratio in % ^a	Deep-sequencing mt:wt ratio in % ^a	Cobas ^b
1	A, B	53/f	V600E	24; 25	21; 21	+
2	A, B	47/f	V600E	25; 26	19; 22	+
3	A, B, C	40/m	V600E	16; 20; 20	9; 15; 14	+
4	A, B	79/m	V600E	53; 59	52; 60	
5	A	55/m	–	–	–	
6	A, B, C	69/m	–	–	–	–
7	A	80/m	V600E	33	33	
8	A, B, C, D, E	53/f	V600E	36; 18; 26; 25; 35	35; 14; 20; 22; 35	
9	A B	87/f	–	–	2 ^{V600E} –	
10	A, C, B	80/m	V600E	56; 62; 45	55; 62; 43	
11	A, B, C	82/f	–	–	1; 1; 1 ^{V600E}	–
12	A	83/m	–	–	–	
13	A, B, C, D	56/m	–	–	–	
14	A, B, C, D, E	57/m	VKS600_602>DT	33; 23; 37; 24; 35	33; 22; 38; 22; 37	
15	A B	74/f	–	9 ^{V600E} –	4 ^{V600E} –	–
16	A, B	65/m	V600E	21; 24	11; 17	

(continued)

Table 1
(continued)

Case	Sample	Age/sex	Sanger sequencing	Pyrosequencing mt:wt ratio in % ^a	Deep-sequencing mt:wt ratio in % ^a	Cobas ^b
17	A B	52/m	– V600K	10 17	7 16	–
18	A	30/m	V600E	28	24	
19	A	75/f	–	11 ^{V600E}	5 ^{V600E}	–
20	D A, B, C, E	66/m	V600E –	22 17; 13; 13; 11	14 7; 6; 6; 5	–
21	A, B	73/m	V600E2	27; 34	31; 36	–
22	A, B	37/m	V600K	39; 39	44; 43	
23	A	71/f	–	–	–	
24	A, B	52/m	–	20; 9 ^{V600E}	9; 3 ^{V600E}	
25	A	54/f	V600E	16	11	
26	A, B, C, D, E, F G, H	66/m	–	–	– 2; 2 ^{V600E}	
27	A, B, C, D, E	54/m	V600K	49; 43; 47; 42; 56	49; 45; 46; 47; 61	+
28	A, B	78/f	V600E	21; 26	9; 12	
29	A, B	44/m	V600E;K601I	61; 39	61; 40	–

Legend: Different samples of the same tumor are specified by A, B, etc.; age in years, *f* female, *m* male

^awt—wild type, mt—mutant

^b+, “–” Mutation detected, “–” Mutation not detected according to the cobas® 4800 report (Roche)

the unambiguous identification of *BRAF* variants in case of mutations beyond V600E/K/D/R/A.

A commercially available Pyrosequencing assay for *BRAF* state detection—therascreen® BRAF Pyro® Kit (Qiagen)—is designed to analyze the antisense strand of *BRAF* starting directly at codon V600. In this particular case, due to mismatching of sequencing primer, a sample with variant mutations downstream from codon V600 will be falsely identified as wild type. Moreover, V600K or V600R mutants may be interpreted as a false V600E mutation at mutant-to-wild-type ratio equal to 25 % or less.

Therefore, we designed a Pyrosequencing assay U-BRAF^{V600} (Patent No. 12153477.0-1404) analyzing the sense strand of human *BRAF* within the activation segment in exon 15 towards the mutations, deletions, and/or insertions, which affect the codons downstream from V600.

We designed the novel dispensation order U-BRAF^{V600}—G₁T₂A₃C₄A₅C₆G₇(A₈T₉)[A₁₀C₁₁T₁₂]G₁₃A₁₄T₁₅C₁₆T₁₇[A₁₈G₁₉] [13]. Because the knowledge of specific variant in each case could explain the altered Pyrogram® tracing created by a change in order and/or quantity of incorporation of each nucleotide, we embedded the two recognition patterns [A₁₀C₁₁T₁₂] and [A₁₈G₁₉], enabling the simultaneous identification of hotspot V600E mutation together with variant mutations p.V600E2 (c.TG1799_1800AA) and p.V600K (c.GT1798_1799AA), tandem mutation p.[V600E;K601I] (c.TG1799_1800AA;A1802T), and complex in-frame mutation p.VKS600_602>DT (c.TGAAAT1799_1804>ATA) [14]. The presence of variant mutations affects the Pyrogram® sequence pattern by redistribution of nucleotide incorporation in the mutant DNA sequence, resulting in a unique Pyrogram for each BRAF mutation (Fig. 1). Both recognition patterns differentiate the individual mutations by the presence of the corresponding peaks characteristic for each mutation variant. Furthermore, the ratio A₈:T₁₂ distinguishes between mutations V600E2 (5:1) and V600K (3:1) (Table 2).

Importantly, unique recognition patterns embedded into U-BRAF^{V600} make it possible to analyze all five different mutations in a single assay. Moreover, compared to Sanger sequencing, where complex deletions and/or insertions require laborious manual analysis, the complex in-frame mutation p.VKS600_602>DT is easily identified using binary (yes/no) data of recognition patterns (Table 3). All mutant variants were quantitatively detectable by the newly developed U-BRAF^{V600} assay. These results were confirmed by ultra-deep sequencing (see Note 1). In contrast to all other *BRAF* state detection methods, the U-BRAF^{V600} assay is capable of automated quantitative identification of at least 36 previously published *BRAF* mutations. Using a threshold of a minimum of 5 % mutated cells in the background of wild-type cells, U-BRAF^{V600} assay design completely excludes false-negative wild-type results (see Note 2).

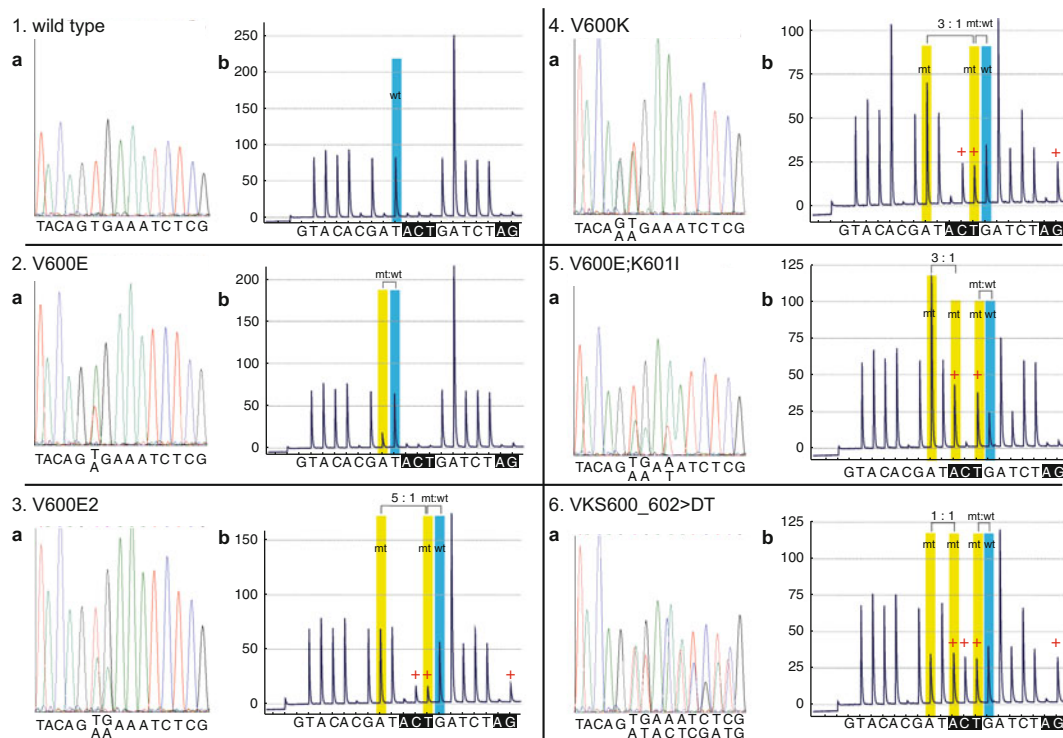


Fig. 1 BRAF mutation analysis by Sanger sequencing and Pyrosequencing-based assay U-BRAF^{V600}. (a) Sanger sequencing; (b) Pyrosequencing-based assay U-BRAF^{V600}. “+” indicates the positive peaks of the dispensation nucleotides within recognition patterns of U-BRAF^{V600} assay. *mt* mutant, *wt* wild type. Recognition patterns are shown in *black boxes*

According to recognition pattern signatures, nine groups were specified as well as four unique mutation variants (Table 3). Importantly, each *BRAF*-mutated variant, including hypothetical one, consists of the features that are unique for each mutation within one group, which enables U-BRAF^{V600} data analysis by the algorithm for *BRAF* state classification (Fig. 2).

In summary, the single-reaction format and data analysis automation make our approach suitable for the assessment of large clinical cohorts. Therefore, we suggest U-BRAF^{V600} assay as a powerful sequencing-based diagnostic tool to automatically identify *BRAF* status.

2 Materials

2.1 Sample Preparation

1. FFPE tissue blocks.
2. Disposable microtome blades (one blade per each FFPE tissue block).
3. Xylene.

Table 2

Dispensation order of five mutated *BRAF* variants detected by U-BRAF^{V600}

BRAF Mutation		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19				
		T	A	C	A		C	G		A	T	A	C	T	G		A		T	C	T	A	G
1	p.V600E (T1799A)	mt wt																					
	wild type	T	A	C	A		G			T			G	A	A	A	T	C	T	-	-		
	c.T1799A	T	A	C	A		-	G		A		-	-	-	G	A	A	A	T	C	T	-	-
2	p.V600E2 (TG1799_1800AA)	mt 5 : 1 mt_wt																					
	wild type	T	A	C	A	G				T			G	A	A	A	T	C	T	-	-		
	c.TG1799_1800AA	T	A	C	A	G	A	A	A	A	A	T	-	C	T	-				C	-	-	G
3	p.V600K	* mt 3 : 1 mt_wt																					
	wild type	T	A	C	A		G			T			G	A	A	A	T	C	T				
	c.GT1798_1799AA	T	A	C	A	A	G	A	A	A	T	-	C	T	-				C	-	-	G	
4	p.V600E;K601I	mt 3 : 1 mt mt_wt																					
	wild type	T	A	C	A		G			T			G	A	A	A	T	C	T				
	c.TG1799_1800AA;A1802T	T	A	C	A		-	G	A	A	A	T	A	-	T	-				C	T	-	-
5	p.VK5600_602>DT	mt wt																					
	wild type	T	A	C	A		G			T			G	A	A	A	T	C	T				
	c.TGAAAT1799_1804>ATA	T	A	C	A		-	G		A		T	A	C	T	-				C	-	-	G

* AS=1Awt + 3Amt

Recognition patterns are indicated in black, individual mutation features are marked in grey
dispensation order's nucleotides, which are involved into mt:wt ratio, are highlighted in bold.

- Ethanol (96–100 %).
- Tissue lysis buffer ATL (Qiagen).
- Proteinase K.
- Lysis buffer AL (Qiagen).
- Automated DNA Extractor Qiasymphony (Qiagen). Alternatively, QIAamp FFPE DNA kit (Qiagen).
- Qubit® Fluorometer and Qubit® dsDNA HS Assay (Thermo Fisher Scientific).

2.2 PCR Amplification

- Custom primers for PCR amplification of *BRAF* exon 15:
 - U-BRAF-F (forward primer):
5'-ATGCTTGCTCTGA TAGGAATG-3'.
 - U-BRAF-R (biotinylated reverse primer):
Biotin-TEG-5'-AGCATCTCAGGGCCAAAAAT-3'.
- Conventional Taq DNA polymerase, e.g., BioTherm Taq DNA polymerase (Genecraft).
- 1× TBE buffer.
- LE Agarose.
- SubCell electrophoresis unit.
- 1× GelRed solution (Biotium).

Table 3
Recognition patterns for 36 *BRAF* mutations by U-BRAF^{V600} assay

Mutation	Recognition patterns						Unique properties of each mutation within one group	mt:wt ratio ^a	COSMIC database ^b
	C6	A10	C11	T12	A18	G19			
1									
p.V600E(1)	-	-	-	-	-	-	A8 = Amt; T9 = Twt	A8 ^{mt} ; T9 ^{wt}	COSM476 [12]
p.T599del	-	-	-	-	-	-	Absence of A8; absence of mutant T2, C4, and A5	[I3-I5] ^{mt} ; A5 ^{wt}	- [15]
p.V600L	-	-	-	-	-	-	Absence of A8; G7 = Gwt; T9 = [Twt + 2Tmt]	[I13-I7] ^{mt} ; G7 ^{wt}	COSM33808 [16]
p.V600M	-	-	-	-	-	-	Absence of A8; G7 = Gwt; T9 = [Twt + Tmt]	[I13-I7] ^{mt} ; G7 ^{wt}	COSM1130 [17]
p.V600R(2)	-	-	-	-	-	-	A5 = Awt; G13 = [Gwt + 2Gmt]	A8 ^{mt} ; T9 ^{wt}	COSM1127 [18]
p.K601E	-	-	-	-	-	-	Absence of A8; G13 = [Gwt + 2Gmt]; A14 = [3Awt + 2Amt]	[I13-I7] ^{mt} ; [2I7-I13] ^{wt}	COSM478 [19]
p.K601N	-	-	-	-	-	-	Absence of A8; T9 = [Twt + Tmt]; A14 = [3Awt + 2Amt]	[I15-I9] ^{mt} ; [2I9-I15] ^{wt}	COSM1132 [20]
2									
p.V600E;K601I	-	+	-	+	-	-	A8:A10 = 3:1	T12 ^{mt} ; G13 ^{wt}	COSM475 [21]
p.V600D	-	+	-	+	-	-	A8 : A10 = 1 : 3	T12 ^{mt} ; G13 ^{wt}	COSM26491
p.V600G	-	+	-	+	-	-	G7 = Gwt + 3Gmt; absence of A8; A10 = 3Amt; G13 = Gwt	T12 ^{mt} ; G13 ^{wt}	COSM477 [12]
3									
p.V600E(2)	-	-	+	+	-	+	A8:T12 = 5:1	T12 ^{mt} ; G13 ^{wt}	COSM475 [21]
p.V600K	-	-	+	+	-	+	A5 = [Awt + 3Amt]; A8:T12 = 3:1	T12 ^{mt} ; G13 ^{wt}	COSM473 [12]
p.V600R(1)	-	-	+	+	-	+	A5 = [Awt + 2Amt]; G7 = [Gwt + 2Gmt]; A8:T12 = 3:1	T12 ^{mt} ; G13 ^{wt}	COSM474 [12]
p.V600_K601>E	-	-	+	+	-	+	A8 : T12 = 2 : 1	T12 ^{mt} ; G13 ^{wt}	COSM1133 [22]
p.TVKS599_603>I	-	-	+	+	-	+	Absence of mutant C4, A5, G7; absence of A8	C11 ^{mt} ; G7 ^{wt}	COSM30605 [23]

4	p.T599T;V600E p.T599_V600>RE p.K601R p.K601K	- - - -	- - - -	- - - -	- - - -	- - - -	+	Absence of mutant A5, G7; absence of A8 Absence of mutant C4, A5; A14 = [3Awt + Amt] Absence of A8; A14 = [3Awt + Amt] Absence of A8; A14 = [3Awt + 2Amt]	G19 ^{mt} ; G7 ^{wt} A8 ^{mt} ; T9 ^{wt} G19 ^{mt} ; T15 ^{wt} G19 ^{mt} ; T15 ^{wt}	COSM24963 [24] COSM476 - [25] COSM13625 [17] COSM28507 [26]
5	p.K601Q p.VKSRWS600_605>D p.VKSRWS600_605>EK	- - -	- - -	- - -	- - -	- - -	- - -	Absence of A8; T9 = [Twt + Tmt]; A18 = 2Amt A8 = Amt; G13 = [Gwt + 3Gmt]; C16 = [Cwt + 3Cmt] A8 = 4Amt; G13 = [Gwt + 4Gmt]; C16 = [Cwt + 3Cmt]	1/2 A18 ^{mt} ; T15 ^{wt} A18 ^{mt} ; T17 ^{wt} A18 ^{mt} ; T17 ^{wt}	COSM1066665 [27] COSM1129 [28] COSM306133 [29]
6	p.V600K;S602S p.T599A	- -	- -	+	- -	- -	+	A8 = 3Amt Absence of A8	T17 ^{mt} ; C11 ^{wt} C11 ^{mt} ; T9 ^{wt}	COSM473 [24] COSM21611 - [30]
7	p.T599_V600insT(1) p.T599_V600insV p.V600>YM	- - -	+	+	- - -	- - -	+	Absence of mutant G7; T9 = [Twt + Tmt] Absence of mutant A3, C4; T9 = [Twt + 2Tmt] Absence of A8; absence of mutant G7, G13, and T17	C11 ^{mt} ; C16 ^{wt} A10 ^{mt} ; A5 ^{wt} C11 ^{mt} ; G13 ^{wt}	COSM30730 [31] COSM21616 [32] - [33]
8	p.T599I;V600E p.A598_T599insKKGNFGLA p.T599_V600>IAL	- - -	+	- -	- -	- -	+	A14 = 3Awt + Amt; T17 = Twt A3 = Awt + 7Amt; absence of mutant A14 Absence of mutant A14; T17 = [Twt + 3Tmt]	A10 ^{mt} ; G7 ^{wt} A10 ^{mt} ; G7 ^{wt} A10 ^{mt} ; G7 ^{wt}	COSM472 [34] COSM476 - [25] COSM33780 [35]

(continued)

Mutation	Recognition patterns						Unique properties of each mutation within one group	mt:wt ratio ^a	COSMIC database ^b
	C6	A10	C11	T12	A18	G19			
9 p.T599_V600insTT	-	+	+	+	-	+	Absence of A8; absence of mutant A5; G19= 2Gmt	T12 ^{mt} ; T15 ^{wt}	COSM26459 [18]
p.VKS600_602>DT	-	+	+	+	-	+	A8= Amt; absence of mutant A14	T12 ^{mt} ; T15 ^{wt}	COSM1162153 [36]
10 p.T599_V600insT(2)	+	-	-	-	-	+	Unique; C4= A8	C11 ^{mt} ; C16 ^{wt}	COSM36245 [37]
11 p.T599_V600insDFLAGT	-	-	-	-	+	+	Unique; T9= [Twt + 4Tmt]	A18 ^{mt} ; 1/3C11 ^{wt}	COSM26504 [22]
12 p.V600A	-	-	+	-	-	-	Unique; absence of A8	C11 ^{mt} ; T9 ^{wt}	COSM18443 [17]
13 p.VKSRRWS600_605>DV	-	-	-	-	-	+	G19=3Gmt; absence of mutant A14; T15= [Twt + 2Tmt]	A8 ^{mt} ; T17 ^{wt}	COSM33764 [38]

^awt—wild type, mt—mutant; I—intensity value of correspondent nucleotide dispensation. A-peak reduction factor 0.9 should be taken into consideration

^bCatalogue of Somatic Mutations in Cancer (COSMIC) database, version 62 (Wellcome Trust Sanger Institute)

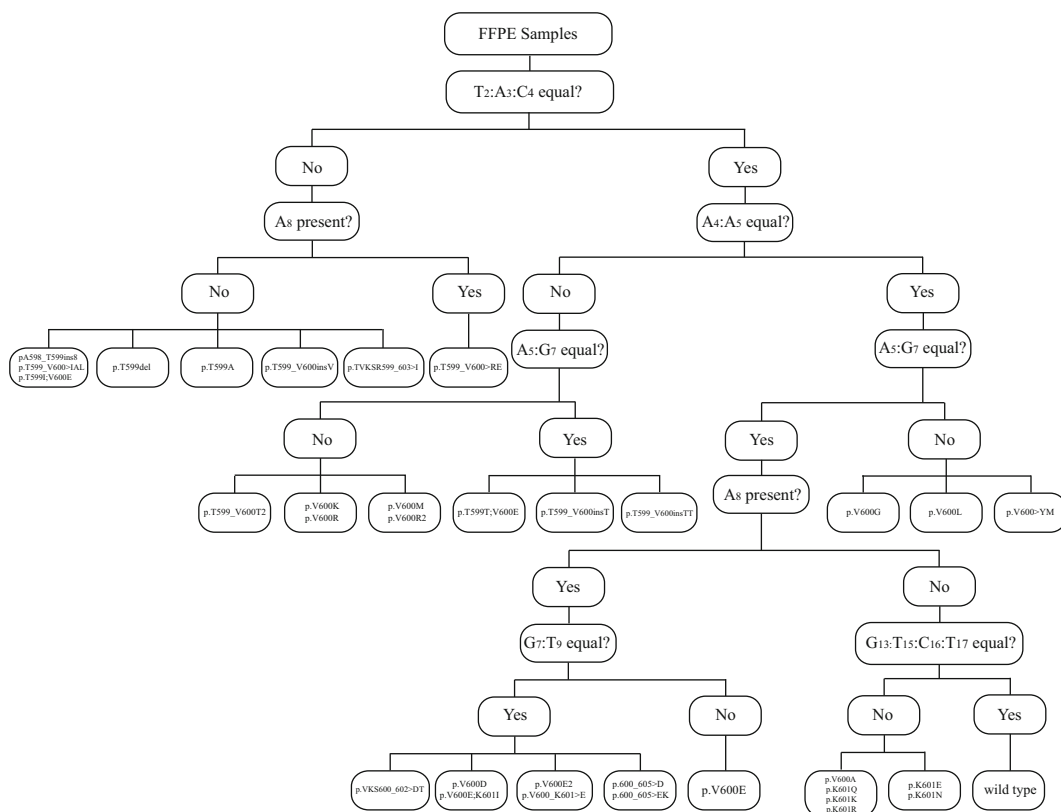


Fig. 2 Algorithm for automated BRAF state classification of U-BRAF^{V600} Pyrosequencing data analysis. Reduction factors for both A-peak and dispensation steps should be taken into consideration calculating individual peak intensities

2.3 Pyrosequencing Reaction

All reagents used for this step should be stored at room temperature except for the streptavidin-coated beads and Pyrosequencing reagents (+4 °C) as well as the Pyrosequencing primer (−20 °C).

1. Custom primer for Pyrosequencing U-BRAF-600-Pyroseq:
5'-AATAGGTGATTTTGGTCTAGC-3'.
2. Streptavidin Sepharose High Performance (GE Healthcare).
3. Pyrosequencer PyroMark® Q24 or PyroMark Q96 (Qiagen).
4. PyroMark Gold Q96 Reagents (Qiagen).
5. Binding buffer: 10 mM Tris-HCl, pH 7.6, 2 M NaCl, 1 mM EDTA, 0.1 % Tween-20.
6. 70 % Ethanol.
7. Denaturation solution: 0.2 M NaOH.
8. Washing buffer: 10 mM Tris-acetate, pH 7.6, adjusted with 4 M glacial acetic acid.
9. Annealing buffer: 20 mM Tris-acetate, pH 7.6, 2 mM magnesium acetate.
10. Deionized water.

3 Methods

The protocol for U-BRAF analysis can be subdivided into four steps:

- *DNA extraction* from FFPE tissues.
- *PCR amplification* of genomic DNA fragment containing the complete exon 15 of *BRAF*.
- The *Pyrosequencing reaction* of amplified fragments.
- *BRAF Mutation State Detection*.

3.1 DNA Extraction

1. Prepare seven 10 μm thick unstained sections from one FFPE tissue block in a 2 mL Eppendorf tube using a new disposable microtome blade for each FFPE tissue block. Avoid cross-contamination by wiping down the knife holder and anti-roll plate with 100 % ethanol in between each block.
2. Extract paraffin from unstained sections by adding 1 mL xylene followed by 1 mL ethanol (96–100 %) according to the manufacturer's instructions (QIAamp FFPE DNA kit).
3. Resuspend FFPE sample in 180 μL ATL buffer and 20 μL Proteinase K.
4. Incubate at 56 °C overnight.
5. Extract total genomic DNA according to the manufacturer's instructions, using an automated DNA extractor (e.g., Qiasymphony™) or the QIAamp FFPE DNA kit.
6. Quantify the eluted DNA using a Qubit® dsDNA HS Assay and Qubit® fluorometer.
7. Store at –20 °C until use.

3.2 PCR Amplification

A strong and specific PCR product is required for a successful Pyrosequencing reaction (*see Note 3*). For the Pyrosequencing assay, the region of human *BRAF* spanning mutation sites within the activation segment in exon 15 was amplified using forward primer U-BRAF-F and biotinylated reverse primer U-BRAF-R.

1. Prepare 50 μL PCR mix by adding:

5 μL	10× Biotherm buffer (contains 1.5 mM MgSO_4)
5 μL	dNTP mix (2.5 mM stock solution)
0.1 μL	U-BRAF-F primer (100 pmol/ μL)
0.1 μL	U-BRAF-R primer (100 pmol/ μL)
0.5 μL	BioTherm Taq DNA polymerase (5 U/ μL)
$x\mu\text{L}$	Genomic DNA (2–10 ng)
$xx\mu\text{L}$	ddH ₂ O

- Run PCR amplification using the following protocol:

Denaturation step:	98 °C—2 min
Cycling (35 cycles):	98 °C—10 s
	54 °C—30 s
	72 °C—20 s
Extension step:	72 °C—10 min

- Analyze the amplified fragments by running 8 µL PCR mix on a 2 % agarose gel in 1× TBE buffer at 8 V/cm using a SubCell electrophoresis unit.
- Stain the 2 % agarose gel in 1× GelRed solution for 30 min.
- Verify the successful PCR amplification by visualizing the 229 bp long fragment under UV light.
- Store amplified PCR products at –20 °C until use.

3.3 Pyrosequencing Reaction

The Pyrosequencing procedure is performed to identify mutations at codons T599 to S602 (5'-TACAGTGAAATCT-3') using the sequencing primer U-BRAF-600-Pyroseq. 20 µL PCR product (400–500 ng) was used for Pyrosequencing according to the manufacturer's instructions (Pyromark Q24).

- Set up new AQ assay using Pyromark Q24/Q96 software as follows:
 - Add "TACAGWGAAATCTAGT" into the field "Sequence to Analyze."
 - Press "Generate Dispensation Order"-button.
 - Edit the "Dispensation Order" to obtain "GTACACGAT ACTGATCTAG."
 - Apply this AQ assay to the designed run template.
- Prepare a Pyrosequencing plate with U-BRAF-600-Pyroseq primer (*see Note 3*) and prepare according to the manufacturer's instructions (Qiagen). A detailed description of the template preparation can be found in other chapters of this volume, e.g., Chapter 7.
- Run the Pyrosequencer according to the manufacturer's instructions (Qiagen).

3.4 BRAF State Detection

The dispensation order of the U-BRAF^{V600} assay is suitable for the identification of a total of 36 previously published BRAF mutation variants affecting codons from T599 to S605 within the activation segment. According to the recognition pattern signatures, we specified nine groups as well as four unique mutation variants (Table 3).

1. Export Pyrosequencing runs as raw data TSV/CSV files.
2. Prepare a spreadsheet file using MS Excel as shown in Table 4.
3. Convert individual TSV/CSV files into Excel files.
4. Collect raw Pyrosequencing data into the field “Dispensation Order” of Excel spreadsheet file created in **step 2**.
5. Use the following operators (*see Note 4*):

Recognition patterns	
C6	=IF(“C6”>5;“+”;“-”)
A10	=IF((“A10”-“G1”)>5;“+”;“-”)
C11	=IF((“C11”-“G1”)>5;“+”;“-”)
T12	=IF((“T12”-“G1”)>5;“+”;“-”)
A18	=IF(“A18”>5;“+”;“-”)
G19	=IF(“G19”>9;“+”;“-”)
Mutation (optional)	
V600E/WT	=IF(“mt:wt ratio”>7;“V600E”;“WT”)
V600K/V600E2	=IF(“A8”/“T12”>4.5;“V600E2”;“V600K”)
Other mutations	

6. Align difficult (non-identified) samples with the corresponding “Recognition patterns” field shown in Table 2 (*see Note 4*).

4 Notes

1. To prove both the sensitivity and the specificity of U-BRAF^{V600} assay, several FFPE samples, which yielded at least 125 ng DNA in 25 µL, were subjected to the cobas® BRAF V600 Mutation Test. In our study, due to initially low biopsy amount, only a few FFPE samples were suitable to perform at least one cobas® BRAF V600 Mutation Test assay analysis [13]. As expected, mutations p.V600E2 (case 21), p.[V600E;K601I] (case 29), and p.VKS600_602>DT (case 14) were not detected by the cobas® BRAF V600 Mutation Test, whereas both p. V600E (cases 1, 2, 3) and p.V600K (case 27) were identified as V600-mutated cases (Table 1). Cases 15, 17, 19, and 20 with low-abundance V600E mutation were not identified by the cobas® 4800 BRAF V600 Mutation Test (Table 1). Therefore, the examined cases were further subjected to ultra-deep-sequencing analysis [13].

Table 4
Spreadsheet for *BRAF* state detection by U-BRAF^{V600}

No. Sample	Mutation	mt:wt ratio ^a	Recognition patterns ^b												Dispensation order												
			C6	A10	C11	T12	A18	G19	G1	T2	A3	C4	A6	C6	G7	A8	T9	A10	C11	T12	G13	A14	T15	C16	T17	A18	G19
1 1A	V600E	24	-	-	-	-	-	-	1.18	95.68	104.9	93.86	105.89	0.59	94.56	27.01	75.8	3.58	4.56	4.3	86.42	257.58	81.15	84.37	78.61	2.61	6.37
2 1B	V600E	25	-	-	-	-	-	-	1.43	62.59	69.77	63.17	69.32	0.64	60.34	18.88	49.89	2.24	3.38	3.18	55.97	169.19	54.29	55.38	53.43	1.8	4.72
8 4A	V600E	53	-	-	-	-	-	-	1.23	59.18	66.44	60.73	65.28	0.59	58.94	35.26	27.95	2.01	1.68	2.62	55.23	166.54	51.34	54.24	51.65	2.06	3.34
9 4B	V600E	59	-	-	-	-	-	-	1.16	67.33	72.84	65.28	71.19	0.5	65.69	43.66	27.82	3.23	1.86	5.09	60.94	182.55	57.89	61.78	58.49	3.23	3.48
11 6A	WT	4	-	-	-	-	-	-	1.31	87.85	95.54	84.35	93.5	0.81	80.64	3.37	80.79	2.75	2.44	2.09	79.84	241.07	74.25	77.09	73.92	2	4.08
12 6B	WT	4	-	-	-	-	-	-	0.96	70.72	81.74	69.65	77.94	0.52	68.61	3	69.32	2.98	1.85	1.83	65.87	204.39	62.81	64.17	61.88	1.84	3.46
13 6C	WT	3	-	-	-	-	-	-	1.24	96.69	106	93.5	103.14	0.74	88.4	3.3	88.39	2.8	2.38	2.31	87.16	266.26	84.27	84.19	82.59	2.21	4.7
33 14A	VKS>DT	33	-	+	+	+	-	+	1.47	111.3	118.5	110.6	115.04	0.81	108.02	40.21	103.59	37.51	35.38	34.48	68.63	199.43	65.01	94.25	60.84	2.54	36.72
37 14E	VKS>DT	35	-	+	+	+	-	+	1.49	91.2	100.4	91.72	96.77	0.83	90.32	36.97	84.26	32.78	32.15	30.8	56.14	161.59	51.59	78.79	49.28	2.47	33.78
51 21A	V600E2	27	-	-	+	+	-	+	1.75	81.03	89.5	78.42	86.04	0.95	79.47	99.69	74.31	3.77	20.91	20.55	54.95	161.01	53.37	70.18	51.29	2.4	21.4
52 21B	V600E2	34	-	-	+	+	-	+	1.32	73.55	81.17	74.4	78.06	0.48	73.22	112.83	67.97	3.89	24.45	24.2	46.45	135.42	44.93	65.78	45.13	2.98	23.45
53 22A	V600K	39	-	-	+	+	-	+	1.67	77.91	86.48	76.2	142.98	0.91	76.31	97.37	70.04	2.31	29.69	28.8	45.34	125.11	41.57	66.77	39.63	1.84	31.67
54 22B	V600K	39	-	-	+	+	-	+	2.71	86.2	95.07	87.51	159.07	0.73	84.74	105.78	78.76	3.69	34.41	32.45	49.91	139.58	44.33	75.41	45.63	1.52	34.74
74 29A	V600E;K601I	61	-	+	-	+	-	-	1.41	72.23	93.06	75.05	82.17	0.95	76.36	134.48	71.39	45.65	7.47	48.49	31.53	97.31	30.5	72.37	74.01	4.48	3.12
75 29B	V600E;K601I	39	-	+	-	+	-	-	1.7	104	116.1	101.7	112.57	1.72	99.88	118.34	95.47	40.88	4.96	38.2	60.55	182.87	63.42	89.91	93.12	3.55	5.02

(continued)

Table 4
(continued)

No. Sample	Mutation	Recognition patterns ^a										Dispensation order																
		mt:wt ratio ^a	C6	A10	C11	T12	A18	G19	G1	T2	A3	C4	A6	C6	G7	A8	T9	A10	C11	T12	G13	A14	T15	C16	T17	A18	G19	
76	A549	WT	7	-	-	-	-	-	-	1.08	36.92	43.43	34.35	40.7	0.75	34.78	2.89	34.63	0.83	1.15	1.14	33.68	105.32	34.07	31.76	32.78	0.79	1.45
77	HeLa	WT	6	-	-	-	-	-	-	3.01	88.58	103.4	93.2	99.68	0.91	91.14	6.47	88.41	3.55	3.54	3.04	86.97	260.18	78.61	80.43	77.51	2.93	6.13
78	WT (cloned)	WT	4	-	-	-	-	-	-	2.92	91.3	105.6	92.75	98.82	1.25	91.04	4.69	91.77	3.23	5.06	4.58	87.54	260.07	82.84	83.38	80.78	4.13	6.52

^a[(0.9 × A8) × 100 / (0.9 × A8 + T9)] for WT and V600E (0.9 – reduction factor for A-peak intensity), [T12 × 100 / (T12 + G13)] for all other mutations

^bOperators:

C6 = IF("C6">5;"+";"-"")
A10 = IF(("A10"-"G1")>5;"+";"-"")
C11 = IF(("C11"-"G1")>5;"+";"-"")
T12 = IF(("T12"-"G1")>5;"+";"-"")
A18 = IF("A18">5;"+";"-"")
G19 = IF("G19">9;"+";"-"")
V600E/WT = IF("mt:wt ratio">7;"V600E";"WT")
V600K/V600E2 = IF("A8"/"T12">4.5;"V600E2";"V600K")

2. In case of low-copy-number BRAF-mutated samples (5 % or less), the recognition patterns can be masked by background noise and, therefore, Pyrograms of V600K, V600E2, or [V600E;K601I] could be very difficult to distinguish from V600E mutation in analyzing only the conventional A₈:T₉ ratio. To simulate low-abundance *BRAF* mutation templates, we subcloned these mutant variants as well as wild-type *BRAF* exon 15. The clones containing V600E, V600E2, or V600K were individually mixed together with the plasmid, containing wild-type *BRAF*, in a proportion from 1 to 10 % mutant variant and subjected to PCR amplification followed by U-BRAF^{V600} Pyrosequencing. Analyzing only the A₈:T₉ ratio, 2 % V600E2 can be misinterpreted either as 10 % V600E or as 4 % V600K (Fig. 3c). In this case, the ratios A₃:A₅, T₉:G₁₃, and T₁₅:C₁₆ should be taken into consideration to estimate the mutant-specific portion in signal intensities of A₅, G₁₃, or C₁₆ (Fig. 3b). In general, the presence of mutations beyond V600E can be determined by the difference in peak intensity values in comparison with correspondent wild-type reference peaks (Figs. 1 and 3c). Importantly, G₁₉ is prone to higher background noise (Table 4) and should therefore be excluded from the low-abundance *BRAF* mutation analysis.
3. We found that at least 400 ng PCR product is required for successful analysis by U-BRAF^{V600} assay, although in this case the signal intensity is constantly reduced by each dispensation step (Fig. 3a). In our study, up to 1 % reduction was observed per dispensation step from the initial intensity value of dispensation nucleotide T₂ resulting in the formula [“reduction factor” × N]%, where “N” is dispensation nucleotide’s number. Therefore, this reduction factor should be taken into consideration in calculating both mutant-to-wild-type ratio and reference peaks’ intensities.
4. Wild-type threshold should be determined according to A₈:T₉ ratio of wild-type reference controls (e.g., A549 or wild-type HeLa cell lines).

Acknowledgments

This work was supported by Deutsche Krebshilfe (DKH108856).

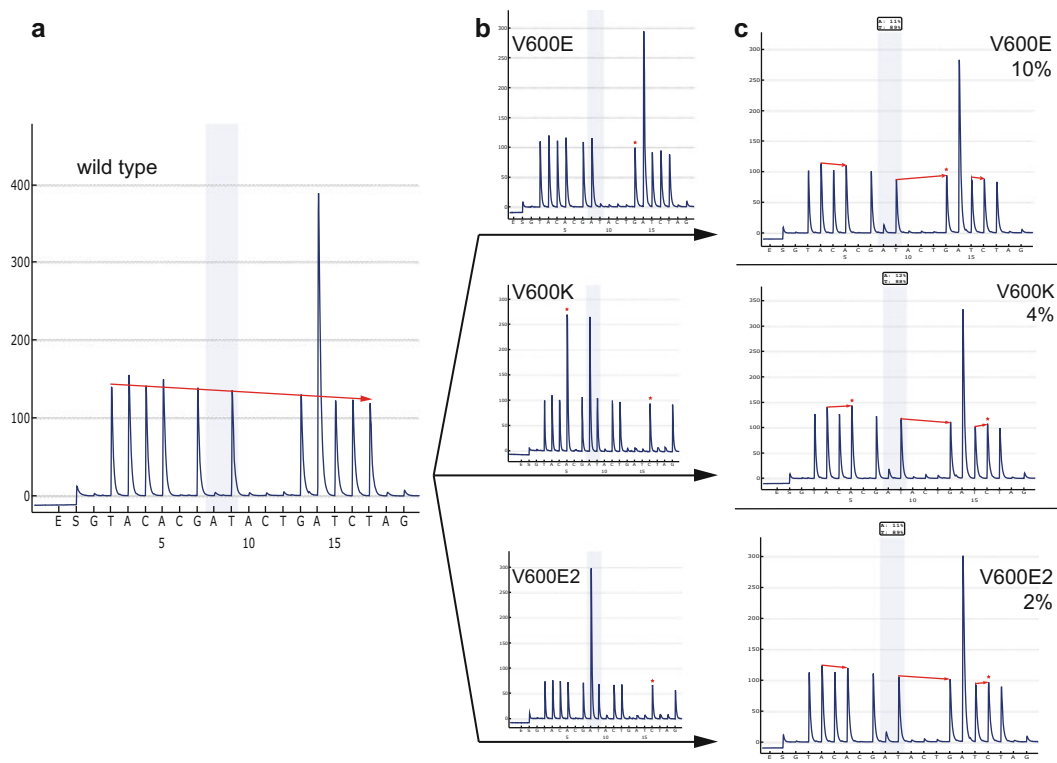


Fig. 3 Low-abundance BRAF mutations. **(a)** Pyrogram of cloned wild-type BRAF. *Red arrow* indicates the reduction of peak intensity values; **(b)** Pyrograms of cloned BRAF mutants. *Red asterisks* indicate the dispensation nucleotide's peaks, which are characteristic for corresponding BRAF mutant in low-copy-number analysis; **(c)** Pyrograms of premixed BRAF mutants with wild type. *Red arrows* indicate the tendency of peak pairs' difference included in low-copy-number analysis. *Red asterisks* indicate the peaks with the contribution of correspondent mutant nucleotides shown in **(b)** (Color figure online)

References

1. Balch CM, Gershenwald JE, Soong SJ et al (2009) Final version of 2009 AJCC melanoma staging and classification. *J Clin Oncol* 27: 6199–6206
2. Lee JH, Choi JW, Kim YS (2011) Frequencies of BRAF and NRAS mutations are different in histological types and sites of origin of cutaneous melanoma: a meta-analysis. *Br J Dermatol* 164:776–784
3. Miller AJ, Mihm MC Jr (2006) Melanoma. *N Engl J Med* 355:51–65
4. Ellerhorst JA, Greene VR, Ekmekcioglu CL et al (2011) Clinical correlates of NRAS and BRAF mutations in primary human melanoma. *Clin Cancer Res* 17:229–235
5. Davies H, Bignell GR, Cox C et al (2002) Mutations of the BRAF gene in human cancer. *Nature* 417:949–954
6. Garnett MJ, Marais R (2004) Guilty as charged: B-Raf is a human oncogene. *Cancer Cell* 6:313–319
7. Safaei Ardekani G, Jafarnejad SM, Tan L et al (2012) The prognostic value of BRAF mutation in colorectal cancer and melanoma: a systematic review and meta-analysis. *PLoS One* 7:e47054

8. Chapman PB, Hausschild A, Robert C et al (2011) Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N Engl J Med* 364:2507–2516
9. Kim SK, Kim DL, Han HS et al (2008) Pyrosequencing analysis for detection of a BRAFV600E mutation in an FNAB specimen of thyroid nodules. *Diagn Mol Pathol* 17: 118–125
10. Heyries KA, Tropini C, Vaninsberghe M et al (2011) Megapixel digital PCR. *Nat Methods* 8:649–651
11. Chung KW, Yang SK, Lee GK et al (2006) Detection of BRAFV600E mutation on fine needle aspiration specimens of thyroid nodule refines cyto-pathology diagnosis, especially in BRAF600E mutation-prevalent area. *Clin Endocrinol (Oxf)* 65:660–666
12. Spittle C, Ward MR, Nathanson KL et al (2007) Application of a BRAF pyrosequencing assay for mutation detection and copy number analysis in malignant melanoma. *J Mol Diagn* 9:464–471
13. Skorokhod A, Helmbold P, Brors B et al (2013) Automated universal BRAF state detection within the activation segment in skin metastases by pyrosequencing-based assay U-BRAF(V600). *PLoS One* 8:e59221
14. Skorokhod A, Capper D, von Deimling A et al (2012) Detection of BRAF V600E mutations in skin metastases of malignant melanoma by monoclonal antibody VE1. *J Am Acad Dermatol* 67:488–491
15. Schulten HJ, Salama S, Al-Mansouri Z et al (2012) BRAF mutations in thyroid tumors from an ethnically diverse group. *Hered Cancer Clin Pract* 10:10
16. Boulalas I, Zaravinos A, Delakas D et al (2009) Mutational analysis of the BRAF gene in transitional cell carcinoma of the bladder. *Int J Biol Markers* 24:17–21
17. Lin J, Goto Y, Murata H et al (2011) Polyclonality of BRAF mutations in primary melanoma and the selection of mutant alleles during progression. *Br J Cancer* 104:464–468
18. Omholt K, Platz A, Kanter L et al (2003) NRAS and BRAF mutations arise early during melanoma pathogenesis and are preserved throughout tumor progression. *Clin Cancer Res* 9:6483–6488
19. Trovisco V, Soares P, Soares R et al (2005) A new BRAF gene mutation detected in a case of a solid variant of papillary thyroid carcinoma. *Hum Pathol* 36:694–697
20. Helmke BM, Mollenhauer J, Herold-Mende C et al (2004) BRAF mutations distinguish anorectal from cutaneous melanoma at the molecular level. *Gastroenterology* 127: 1815–1820
21. Indsto JO, Kumar S, Wang L et al (2007) Low prevalence of RAS-RAF-activating mutations in Spitz melanocytic nevi compared with other melanocytic lesions. *J Cutan Pathol* 34: 448–455
22. Hou P, Liu D, Xing M (2007) Functional characterization of the T1799-1801del and A1799-1816ins BRAF mutations in papillary thyroid cancer. *Cell Cycle* 6:377–379
23. Lupi C, Giannini R, Ugolini C et al (2007) Association of BRAF V600E mutation with poor clinicopathological outcomes in 500 consecutive cases of papillary thyroid carcinoma. *J Clin Endocrinol Metab* 92:4085–4090
24. Deichmann M, Thome M, Benner A et al (2005) Preponderance of the oncogenic V599E and V599K mutations in B-raf kinase domain is enhanced in melanoma cutaneous/subcutaneous metastases. *BMC Cancer* 5:58
25. Jung CK, Im SY, Kang YJ et al (2012) Mutational patterns and novel mutations of the BRAF gene in a large cohort of Korean patients with papillary thyroid carcinoma. *Thyroid* 22:791–797
26. Deichmann M, Krah D, Thome M et al (2006) The oncogenic B-raf V599E mutation occurs more frequently in melanomas at sun-protected body sites. *Int J Oncol* 29:139–145
27. Sarkozy A, Carta C, Moretti S et al (2009) Germline BRAF mutations in Noonan, LEOPARD, and cardiofaciocutaneous syndromes: molecular diversity and associated phenotypic spectrum. *Hum Mutat* 30: 695–702
28. Cruz F 3rd, Rubin BP, Wilson D et al (2003) Absence of BRAF and NRAS mutations in uveal melanoma. *Cancer Res* 63:5761–5766
29. Hunter SM, Gorringer KL, Christie M et al (2012) Pre-invasive ovarian mucinous tumors are characterized by CDKN2A and RAS pathway aberrations. *Clin Cancer Res* 18: 5267–5277
30. Satoh T, Smith A, Sarde A et al (2012) B-RAF mutant alleles associated with Langerhans cell histiocytosis, a granulomatous pediatric disease. *PLoS One* 7:e33891
31. Eisenhardt AE, Olbrich H, Röring M et al (2011) Functional characterization of a BRAF insertion mutant associated with pilocytic astrocytoma. *Int J Cancer* 129:2297–2303
32. Carta C, Moretti S, Passeri L et al (2006) Genotyping of an Italian papillary thyroid carcinoma cohort revealed high prevalence of

- BRAF mutations, absence of RAS mutations and allowed the detection of a new mutation of BRAF oncoprotein (BRAF(V599Ins)). Clin Endocrinol (Oxf) 64:105–109
33. Matsuse M, Mitsutake N, Tanimura S et al (2013) Functional characterization of the novel BRAF complex mutation, BRAF(V600delins YM), identified in papillary thyroid carcinoma. Int J Cancer 132:738–743
 34. Gill M, Renwick N, Silvers DN et al (2004) Lack of BRAF mutations in Spitz nevi. J Invest Dermatol 122:1325–1326
 35. Chiose S, Nikiforova M, Zuo H et al (2009) A novel complex BRAF mutation detected in a solid variant of papillary thyroid carcinoma. Endocr Pathol 20:122–126
 36. Thomas NE, Edminston SN, Alexander A et al (2007) Number of nevi and early-life ambient UV exposure are associated with BRAF-mutant melanoma. Cancer Epidemiol Biomarkers Prev 16:991–997
 37. Schindler G, Capper D, Meyer J et al (2011) Analysis of BRAF V600E mutation in 1,320 nervous system tumors reveals high mutation frequencies in pleomorphic xanthoastrocytoma, ganglioglioma and extra-cerebellar pilocytic astrocytoma. Acta Neuropathol 121:397–405
 38. Barzon L, Masi G, Boschin IM et al (2008) Characterization of a novel complex BRAF mutation in a follicular variant papillary thyroid carcinoma. Eur J Endocrinol 159:77–80

Chapter 7

Pyrosequencing®-Based Identification of Low-Frequency Mutations Enriched Through Enhanced-*ice*-COLD-PCR

Alexandre How-Kit and Jörg Tost

Abstract

A number of molecular diagnostic assays have been developed in the last years for mutation detection. Although these methods have become increasingly sensitive, most of them are incompatible with a sequencing-based readout and require prior knowledge of the mutation present in the sample. Consequently, coamplification at low denaturation (COLD)-PCR-based methods have been developed and combine a high analytical sensitivity due to mutation enrichment in the sample with the identification of known or unknown mutations by downstream sequencing experiments. Among these methods, the recently developed Enhanced-*ice*-COLD-PCR appeared as the most powerful method as it outperformed the other COLD-PCR-based methods in terms of the mutation enrichment and due to the simplicity of the experimental setup of the assay. Indeed, E-*ice*-COLD-PCR is very versatile as it can be used on all types of PCR platforms and is applicable to different types of samples including fresh frozen, FFPE, and plasma samples. The technique relies on the incorporation of an LNA containing blocker probe in the PCR reaction followed by selective heteroduplex denaturation enabling amplification of the mutant allele while amplification of the wild-type allele is prevented. Combined with Pyrosequencing®, which is a very quantitative high-resolution sequencing technology, E-*ice*-COLD-PCR can detect and identify mutations with a limit of detection down to 0.01 %.

Key words Pyrosequencing®, Genotyping, Mutation detection, E-*ice*-COLD-PCR, Real-time PCR, Mutation hotspots, *KRAS*

1 Introduction

Over the last 30 years, a number of molecular methods have been developed for the detection of mutations in different regions of interest such as mutation hotspots. It has become clear that one of the main clinical challenges is the capacity to detect mutations with very high sensitivity in all types of clinical samples for diagnosis, the choice of a personalized treatment regimen, and/or the monitoring of the treatment response. Among the available methods, some require a combination of PCR and sequencing technologies while others are generally used without any downstream sequencing experiments such as high-resolution melting (HRM)

and allele-specific PCR (AS-PCR)-based methods [1, 2]. Assays using sequencing-based mutation detection have the advantage to allow the identification of the mutated base pair (in contrast to HRM) and they do not require prior knowledge of the mutation (in contrast to AS-PCR).

Therefore, Sanger sequencing has become the gold standard for mutation detection. However, Sanger sequencing has an average analytical sensitivity of only 20 % for the mutated alleles in a wild-type background [3]. Pyrosequencing® is a more recent quantitative real-time sequencing method that has also frequently been used for mutation detection [4, 5]. Pyrosequencing is based on the presence or absence of the incorporation of a nucleotide during primer extension [6, 7]. In contrast to Sanger sequencing, which relies on the random incorporation of fluorescent ddNTPs during primer extension steps, only one specific nucleotide is present at any time in the Pyrosequencing reaction. Following nucleotide incorporation, the released pyrophosphate (PPi) is used as substrate in combination with adenosine 5' phosphosulfate (APS) by an ATP sulfurylase to produce ATP [4]. The latter is in turn used by luciferase to oxidize luciferin into oxyluciferin resulting in a light emission proportional to the amount of incorporated nucleotide [4]. The limit of detection of Pyrosequencing has been evaluated around 5 % for the mutant allele, which is thus far more sensitive than Sanger sequencing [3].

To further increase the limit of mutation detection, a large variety of PCR-based methods have been developed to enrich for unknown mutations in samples prior to sequencing [8]. For example, coamplification at lower denaturation temperature (COLD- or *full*-COLD) PCR has been developed to allow high enrichment of mutations located in the PCR amplification product (Fig. 1) [9].

The principle of this method is based on the formation of WT-WT homoduplexes and WT-mutant heteroduplexes at 70 °C after an initial denaturation step and followed by selective denaturation of the WT-mutant heteroduplexes at a critical temperature (T_c) determined as the melting temperature (T_m) of the amplicon minus one degree [9]. Then, primer annealing and elongation steps are performed where mutant and WT alleles undergo exponential or linear amplification, respectively, leading to enrichment of the mutated alleles. This protocol has been slightly modified in the *fast*-COLD-PCR assay where no initial denaturation step is performed and no mutant-WT heteroduplexes are formed (Fig. 1). *fast*-COLD-PCR enables a stronger enrichment of mutations compared to *full*-COLD-PCR for T_m -reducing mutations such as G:C>A:T or G:C>T:A, which are the most frequent occurring mutations [10, 11]. However, mutation enrichment is completely abolished for T_m -equivalent and T_m -increasing mutations in *fast*-COLD-PCR [10, 11]. Therefore a new COLD-PCR-based approach has been developed enabling a high enrichment of all

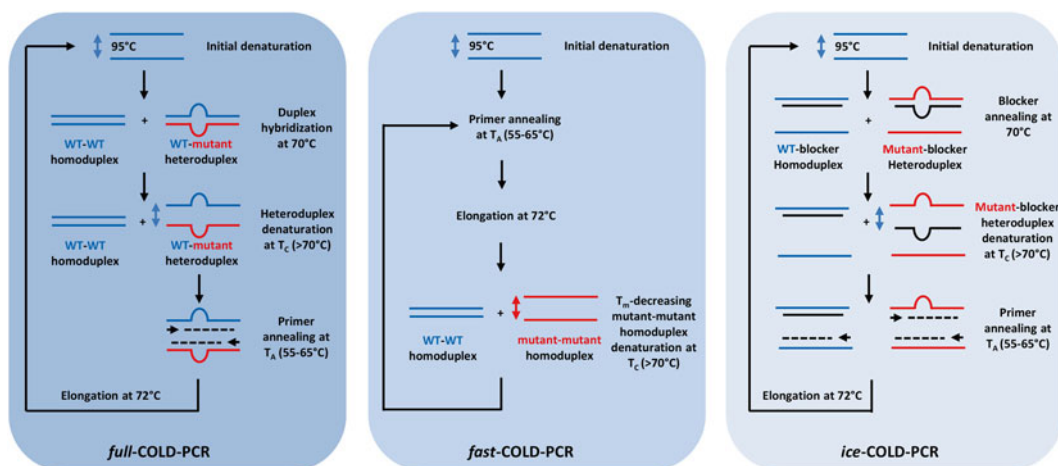


Fig. 1 Principle of COLD-PCR-based methods. In the *full*-COLD-PCR assay, mutation enrichment is performed after an initial denaturation step, followed by the formation of WT-WT homoduplexes and WT-mutant heteroduplexes at 70 °C, selective denaturation of WT-mutant heteroduplexes at the T_c and primer annealing and elongation. In a *fast*-COLD-PCR assay, WT-WT and mutant-mutant homoduplexes are formed after a first denaturation, annealing, and elongation cycle and enrichment of the mutation is then performed during PCR cycles including a selective denaturation of mutant-mutant homoduplexes at the T_c followed by primer annealing and elongation steps. In the *ice*-COLD-PCR assay, WT-blocker homoduplexes and mutant-blocker heteroduplexes are formed after an initial denaturation step followed by the selective denaturation of mutant-blocker heteroduplexes at the T_c and primer annealing and elongation resulting in mutation enrichment. T_c : critical temperature

types of mutations, called *improved* and *complete enrichment* (*ice*)-COLD-PCR [11]. *ice*-COLD-PCR introduces a non-elongable blocker probe that is complementary to the WT sequence and that overlaps with five nucleotides at the 3' end of each primer, which is used to block WT-homo- and mutant heteroduplex formation [11].

One of the drawbacks of COLD-PCR-based methods is the use of a critical temperature (T_c), which has to be very accurate as a slight variation of as little as 0.2 °C could completely abolish the mutation enrichment [11]. Moreover, *ice*-COLD-PCR requires two pairs of overlapping PCR primers to determine the correct T_c as well as a first PCR amplification [11]. Therefore, we developed Enhanced-*ice*-COLD-PCR (E-*ice*-COLD-PCR) as a modified version of *ice*-COLD-PCR which contains chemically modified nucleotides (locked nucleic acid (LNA) bases) in a primer-overlapping blocker probe (Fig. 2) [12].

The determination of the T_c requires only one primer pair and a strong mutation enrichment is achieved, which outperformed the originally reported approach. Efficient enrichment can be achieved over a large range of temperatures making the method easily usable on different thermocycler platforms [12]. Another advantage of E-*ice*-COLD-PCR is that it requires no pre-amplification PCR and

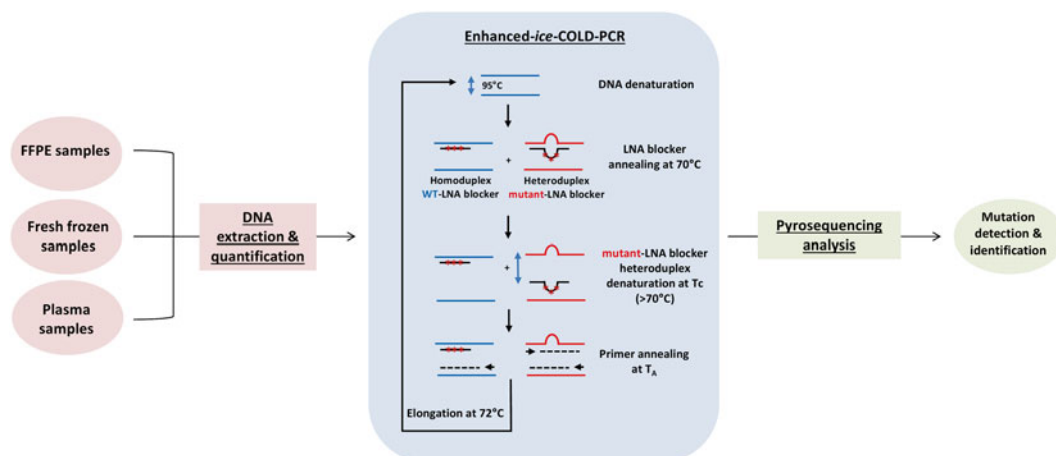


Fig. 2 General workflow for the enrichment, detection and identification of mutations by Enhanced-*ice*-COLD-PCR and Pyrosequencing. Different types of samples can be used for this method including DNA from FFPE, fresh-frozen, and plasma samples. After quality control, an E-*ice*-COLD-PCR reaction is performed using two PCR primers, one of them being biotinylated, flanking the region with the mutation (hotspot) of interest. The blocker probe complementary to the WT sequence allows the enrichment of the mutation during the PCR, performed with a cycling protocol in which only the mutant sequences undergo exponential amplification while the amplification of WT sequences remains linear. The resulting PCR product is purified prior to the Pyrosequencing analysis, which allows the accurate and ultra-sensitive detection and identification of mutations present in the analyzed region of interest

is applicable to different quantities and types of samples rendering this method very robust and versatile [12]. Thus mutation enrichment using E-*ice*-COLD-PCR combined with Pyrosequencing for a sequence-based readout of the enriched alleles is currently the easiest, most rapid, and sensitive sequencing-based mutation detection and identification assay format.

In this chapter, we describe all steps necessary for the development of an Enhanced-*ice*-COLD-PCR assay for the detection and identification of mutations in a region of interest. We have included all steps necessary for the design of the assay, the identification of the optimal experimental conditions, and the analysis and interpretation of Pyrograms[®]. Results are exemplified on the analysis of *KRAS* mutations at the mutation hotspot in codons 12 and 13, which are of relevance for the choice of treatment in a variety of cancers [12].

2 Materials

2.1 Assay Design

1. Design of PCR primers using Primer 3 [13, 14] (<http://bioinfo.ut.ee/primer3-0.4.0/>), verification of primer efficiency (<http://www.premierbiosoft.com/netprimer/>) and of potential secondary structures using mFold [15] (<http://mfold.rna.albany.edu/?q=mfold/DNA-Folding-Form>).

2. Identification of polymorphisms inside primers and nonspecific primer annealing using the BLAT tool (<http://genome.ucsc.edu/cgi-bin/hgBlat?command=start>).
3. Manual design of LNA blocker probe and calculation of the T_m using the Exiqon Prediction Tool (<https://www.exiqon.com/ls/Pages/ExiqonTMPredictionTool.aspx>).
4. Design of Pyrosequencing primers using the commercial PSQ assay design software.

2.2 DNA Extraction and Quantification

1. DNeasy® Blood and Tissue Kit.
2. QIAamp® DNA FFPE Tissue Kit.
3. Chemagic Circulating NA Kit special.
4. Quant-iT™ PicoGreen® dsDNA Assay Kit.

2.3 DNA Standards, Identification, and Preparation

1. Human genomic DNA.
2. Identification of cell lines harboring mutations of interest in COSMIC [16] (<http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/>).
3. Genomic DNA of the cell line to be ordered at ATCC (http://www.lgcstandards-atcc.org/?geo_country=fr) or any cell line DNA repository bank.
4. NanoDrop spectrophotometer.

2.4 PCR Amplification

1. PCR primers.
2. Biotinylated PCR primers.
3. 3' Phosphate-modified LNA blocker probes.
4. HotStar Taq DNA Polymerase.
5. dNTPs.
6. SYTO9.
7. LightCycler® 480 Multiwell Plate 96.
8. LightCycler® 96 Real-Time PCR System.

2.5 Agarose Gel Electrophoresis

1. TBE 10×.
2. UltraPure™ Agarose.
3. 100 bp DNA ladder.
4. 6× DNA Loading Dye Buffer Blue.
5. Gel molds and combs.
6. Electrophoresis apparatus.
7. Gel electrophoresis imaging system.

2.6 Sample Preparation for Pyrosequencing

1. 96-Well Low Profile PCR Plate, Skirted.
2. Streptavidin Sepharose HP beads.
3. PyroMark® Q96 Vacuum Workstation (220 V) [17].

4. PyroMark Q96 Plate Low.
5. PyroMark Q96 Sample Prep Thermoplate Low.
6. PyroMark Q96 Vacuum Prep Trough.
7. Binding buffer: 10 mM Tris-HCl, 2 M NaCl, 1 mM EDTA, 0.1 % Tween 20, pH 7.6.
8. Denaturing solution: 0.2 M NaOH.
9. Washing buffer: 10 mM Tris-acetate, pH 7.6.
10. Annealing buffer: 20 mM Tris-acetate, 2 mM Mg-acetate, pH 7.6.
11. Costar microplate sealing tape.
12. Thermomixer comfort.
13. Pyrosequencing primer.

2.7 Mutation Detection and Identification by Pyrosequencing

1. PyroMark Q96 Cartridge.
2. PyroMark Q96 MD Pyrosequencer.
3. PyroMark MD software.
4. PyroMark Gold Q96 Reagents (5 × 96).

3 Methods

The detection, identification, and quantification of mutations by Enhanced-*ice*-COLD-PCR followed by Pyrosequencing can be divided into seven steps (Fig. 2):

1. The design of the assay, which includes the design of the PCR primers, the LNA blocker probe, the Pyrosequencing primer, and the identification of cell line DNA standards.
2. The validation of the PCR primers including the identification of the best primer combination and their annealing temperature.
3. The validation of the sequencing primer and of the mutation present in the cell line used as standard for the optimization of the assay.
4. The identification of the critical temperature and the optimal concentration of the LNA blocker probe.
5. The generation of a Pyrosequencing assay allowing the identification and the quantification of all mutations of interest.
6. The DNA extraction and accurate quantification from samples of different types such as fresh frozen, FFPE, and plasma.
7. The detection and identification of the mutation by E-*ice*-COLD-PCR followed by Pyrosequencing in the sample of interest.

3.1 DNA Extraction and Quantification from Fresh Frozen, FFPE, and Plasma Samples

For an optimal performance of E-*ice*-COLD-PCR, it is necessary that DNA extractions from different types of samples are performed properly and accurate quantification of the resulting DNA quantity has to be performed before downstream applications.

1. For fresh-frozen samples, extract genomic DNA using the DNeasy Blood and Tissue Kit according to the manufacturer's instructions.
2. For FFPE samples, extract genomic DNA using the QIAamp DNA FFPE Tissue Kit according to the manufacturer's instructions.
3. For plasma samples, extract circulating DNA using the Chemagic NA extraction kit according to the manufacturer's instructions. The volume of reagents can be adjusted according to the starting volume of available plasma.
4. Quantify genomic DNA from fresh-frozen samples using the Quant-iT™ PicoGreen® dsDNA Assay Kit according to the manufacturer's instruction.
5. Quantify DNA from FFPE and plasma samples by a quantitative real-time PCR assay using a dilution series of commercial DNA as standards (*see Note 1*).

3.2 Assay Design

The design of the assay is a critical step for Enhanced-*ice*-COLD-PCR and should be carefully performed to avoid any subsequent problems in the following experimental protocol. At least two pairs of PCR primers per region and two LNA blocker probes per assay should be designed and evaluated.

3.2.1 Identification of the Region of Interest and DNA Standards

1. Enhanced-*ice*-COLD-PCR is a method particularly suitable for the detection and identification of mutations in hotspot regions, where different types of mutation occur on a restricted number of nucleotides. Thus the identification of the region of interest is crucial and should include all mutations of potential interest. Identification of known mutations can be done with the help of mutation databases such as the COSMIC database listing more than one million somatic mutations throughout the genome in different types of cancer samples or cell lines.
2. The region of interest should not be larger than 200 nucleotides and its boundaries should be as much as possible free of mutations or should contain only very-low-frequency mutations. The mutation hotspots can extend from one to a few tens of codons (examples: *KRAS* codons 12 and 13, *BRAF* codon 600 and *NRAS* codon 61).
3. For the development of the assay, identify at least one cell line carrying a mutation located in the region of interest from the literature or a database such as COSMIC.

4. Order the corresponding genomic DNA directly from a cell line bank such as ATCC (*see Note 2*).
5. Upon receipt, determine the concentration of the cell line DNA using the Quant-iT™ PicoGreen® dsDNA Assay Kit or similar double-stranded DNA-specific quantification kits (*see Note 3*).
6. Genotype the genomic DNA of the different cell lines by Pyrosequencing in order to validate the presence of the correct mutations as well as their proportions (*see Subheading 3.3*).
7. Perform serial dilutions of mutant and WT commercial DNA to obtain standards with different mutation fractions, e.g., 50, 10, 5, 1, 0.5, and 0.1 % (*see Note 4*). Genotype the dilutions to assess the linearity and the accuracy of the dilutions (*see Subheading 3.3*).

3.2.2 Design of Amplification Primers

1. Design the PCR primers using the freely available Primer 3 software. The length of the amplification product should optimally be around 100–150 bp to be applicable to all types of samples including fragmented DNA such as cell-free circulating DNA and FFPE samples.
2. Report in brackets the mutation hotspot region of interest in order to include the target region in the proposed assays. Optimal primer size should be around 20 nucleotides and the optimal T_m around 60 °C, which are the default settings in Primer 3.
3. For all the different proposed assays, verify the absence of secondary structures using the mfold program to assess the suitability of the assays for real-time PCR. Evaluate the primer efficiencies using the online software NetPrimer.
4. For each assay, investigate the presence of potential polymorphic positions such as SNPs underlying the annealing sites as well as primer specificity using a BLAT search on the UCSC Genome Bioinformatics portal. Discard the assays where amplification primers anneal to potential polymorphic sites and/or when both primers are highly complementary to several regions of the genome (*see Note 5*).
5. Select at least two assays among the proposed ones.
6. Include a biotin on the 5' end of one primer, depending on the sense of the genotyping/Pyrosequencing assay (*see below*).

3.2.3 Design of Pyrosequencing Primers

1. A Pyrosequencing primer is used to detect, identify, and quantify mutations by genotyping. Design Pyrosequencing primers using the commercial PSQ assay design software.
2. The sequencing primer should not anneal to any polymorphic and/or mutation site and should contain at least two nucleotides

not overlapping with the PCR primers on its 3' end to avoid any unspecific signals or annealing to potential primer dimers.

3. At least the last four bases of the 3' end of the sequencing primers should be unique in the amplification product to avoid the generation of background signals, which could interfere with accurate quantification.
4. If possible, select the sense of the sequencing reaction to avoid the presence of "T" mutations, as the complementary Pyrosequencing reaction will incorporate an "A" nucleotide. Adenosine triphosphates can be used as substrate by the luciferase despite the use of dATP α S and can induce a high background signal [18].
5. Select the sense of the sequencing reaction according to the ease of interpretation of the Pyrograms in one direction or the other around the mutation hotspot.
6. Do not place the Pyrosequencing primer too far from the mutation hotspot region as the PyroMark Gold Q96 Reagents used for genotyping experiments can only sequence 30–40 nucleotides.

3.2.4 Design of Pyrosequencing Assays

Pyrosequencing assays are designed using the PyroMark MS software supplied with the Pyrosequencer for genotyping experiments.

1. To design a genotyping assay, insert the DNA sequence beginning at the base beyond the 3' end of the sequencing primer into the text box "sequence to analyze" in the "simplex entries" tab. The sequence must cover the mutation hotspot region and also include nucleotides, which are known to be not mutated and are used as control nucleotides.
2. Include at least one mutation in the nucleotide sequence using the IUPAC nomenclature for the polymorphism. The "dispensation order" tool generates the order in which the different nucleotides are dispensed.
3. For the development of E-ice-COLD-PCR assays and the setup of the experimental conditions, cell line DNA standards bearing known mutations are used. For these steps, create a genotyping assay including only the one mutation specific for the cell line, which is easily manageable by the software. It allows accurate quantification of the mutation.
4. For the detection, identification, and quantification of unknown mutation in the region of interest, insert the nucleotide sequence for all possible combinations of mutations in a single genotyping assay.
5. The software will indicate in the warnings window that the sequence might be too complicated or is ambiguous to allow

an automated determination or quantification of all possible mutations. If the software indicates that “AQ mode cannot be performed” or that “some genotypes will generate the same sequence patterns and will not be distinguishable,” the input sequence should be simplified. In this case, leave only the most frequent mutations in the sequence to generate the dispensation order, which can be quantified by the software.

6. Add manually to the dispensation sequence the remaining possible mutations, which need to be included for mutation screening. These nucleotides are considered by the software as control nucleotides generating no peaks. If these mutations are present in a sample, the software considers it as a failed sample because the dispensation pattern is uncertain. These samples need to be distinguished from failed samples with no signal.
7. Analyze manually the Pyrograms of “failed samples” to detect and identify the presence of possible mutations. Use raw peak intensities for manual or automated mutation quantification (*see Note 6*).

3.2.5 Design of LNA Blocker Probe

1. The LNA blocker probe inhibits elongation of one WT DNA strand during PCR as its nucleotide sequence is fully complementary to the WT sequence. Design the LNA blocker probe over the possible mutated nucleotides between the two amplification primers and overlap five to six nucleotides with the 3' end of one of the amplification primers.
2. The primer partly complementary to the blocker probe should be the most stable one in order to avoid the formation of primer dimers and/or primer self-dimers. Determine the most stable primer bioinformatically using the NetPrimer software.
3. If possible, center the LNA blocker probe on the mutation hotspot and include some nucleotides, which are known to be not mutated.
4. LNA bases allow increased single-nucleotide discrimination by maximizing the T_m differences between a single-nucleotide mismatch compared to a perfect match, except for G-T mismatches when the G is present on the DNA strand bearing LNA bases [19]. Select the sense (forward or reverse) of the LNA blocker probe to include as few as possible G nucleotides in the region containing the mutation.
5. Preferentially include LNA bases on each nucleotide likely to be mutated in order to generate mismatches. If the potentially mutated sequences include only one nucleotide, include an LNA triplet in the probe centered on the nucleotide of interest. Up to six consecutive LNA bases can be included in the blocker probe if the mutation hotspots are composed of less than ten consecutive nucleotides. If the potentially mutated

nucleotides are spread over several tens of nucleotides, include one LNA base every two, three, or four nucleotides in the blocker probe.

6. At least six nucleotides at the 5' and 3' ends of the blocker probe should be devoid of LNAs.
7. Verify the T_m of the LNA blocker probe bioinformatically using the Exiqon T_m Prediction Tool. The T_m should be comprised between 75 °C and 90 °C. If not, modulate the length of the LNA blocker probe and/or the number of incorporated LNA bases.
8. Add a phosphate residue to the 3' end of the probe to avoid any elongation of the LNA blocker probe during the PCR reaction.

3.3 Determination of the Optimal E-ice-COLD-PCR Conditions

3.3.1 Validation of PCR Primers

For each region of interest at least two PCR primer pairs should be ordered for evaluation and only the very best ones should be selected for further Enhanced-*ice*-COLD-PCR application.

1. The first step of the validation consists of the determination of the optimal annealing temperature for each primer pair. This is performed on a gradient thermocycler using 25 ng of commercial WT genomic DNA as template.
2. Typical PCR reaction conditions are 1× HotStar Taq buffer supplemented with 1.6 mM $MgCl_2$, 200 mM of each dNTP, 1.0 U of HotStar Taq polymerase, and 200 nM of forward and reverse primers in a total volume of 25 µL.
3. Typical PCR cycling conditions include an initial denaturation step performed 10 min at 95 °C, followed by 50 cycles of 30-s denaturation at 95 °C, 20-s of a gradient annealing ranging from 50 to 70 °C, and 10-s elongation at 72 °C. Final extension is performed 5 min at 72 °C and the cooling is performed at 40 °C.
4. Deposit 10 µL of PCR product supplemented with 1× loading dye as well as a DNA ladder sample on a 2 % agarose gel including a fluorescent dye (example: ethidium bromide). Perform a horizontal electrophoresis in TBE 1× at 5 V/cm during 1 h. The specificity of the amplifications is assessed by evaluation of the length of the PCR product and the highest temperature allowing a strong and specific amplification product is selected as the optimal annealing temperature.
5. Evaluate the different PCR primer pairs for a real-time PCR application using the optimal annealing temperature. PCR reaction conditions are the same as described in **step 2** supplemented with 2 µM of SYTO9 as fluorescent dye.
6. PCR cycling conditions include an initial denaturation step for 10 min at 95 °C, followed by 50 cycles of 30-s denaturation at

95 °C, 20 s at the optimal annealing temperature, and 10-s elongation at 72 °C. Acquire fluorescence at the end of the annealing step in each cycle. Final steps include a melting curve performed from 65 °C to 97 °C and cooling to 40 °C (*see Note 7*).

7. Amplification curves of each assay allow the identification of nonspecific products if some signal is present in the no-template reactions. Melting curves allow the determination of the T_m of the PCR amplification product and of primer dimers present in the no-template reactions. Select the assays generating no or late fluorescent signal in the no-template reaction for further application of the E-*ice*-COLD-PCR assays.

3.3.2 Determination of the Critical Temperature, the Optimal Blocker Probe Concentration, and the Limit of Detection of the Assay

The critical temperature (T_c) is the optimal temperature where blocker-mutant heteroduplexes are denatured while blocker-WT homoduplexes remain hybridized. This temperature is essential in E-*ice*-COLD-PCR reaction as it allows an optimal enrichment of the mutation. Figure 3 shows an example of the optimization of the critical temperature using different temperatures and blocker concentrations.

1. Prepare a stock solution for LNA blocker probes at a concentration of 10 μ M. Prepare a 250 nM dilution from the stock solution. Store all solutions at -20 °C.
2. To determine the critical temperature, perform an E-*ice*-COLD-PCR using a T_c gradient, different concentrations of LNA blocker probes, and 25 ng of a DNA standard sample containing a mutation fraction of 1 or 0.5 %. Recommended concentrations for the blocker probes are as follows: 0, 10, 20, 30, and 50 nM (Fig. 3).
3. PCR reaction conditions are 1 $\times$ HotStar Taq buffer supplemented with 1.6 mM $MgCl_2$, 200 mM of each dNTP, 1.0 U of HotStar Taq polymerase, 200 nM of forward and reverse primers, 2 μ M of SYTO9, and different concentrations of blocker probes in a 25 μ L volume.
4. PCR cycling conditions include an initial denaturation step for 10 min at 95 °C, followed by six cycles of standard PCR (30-s denaturation at 95 °C, 20 s at optimal primer annealing temperature, and 10-s elongation at 72 °C) and 44 cycles of E-*ice*-COLD-PCR (20-s denaturation at 95 °C, 30-s blocker annealing at 70 °C, 20 s at T_c using a temperature gradient ranging from 70 °C to 90 °C, 20 s at optimal primer annealing temperature, and 10-s elongation at 72 °C). Acquire fluorescence at the end of the annealing step in each cycle. Final steps include a melting curve performed from 65 °C to 97 °C and cooling to 40 °C (*see Note 8*).

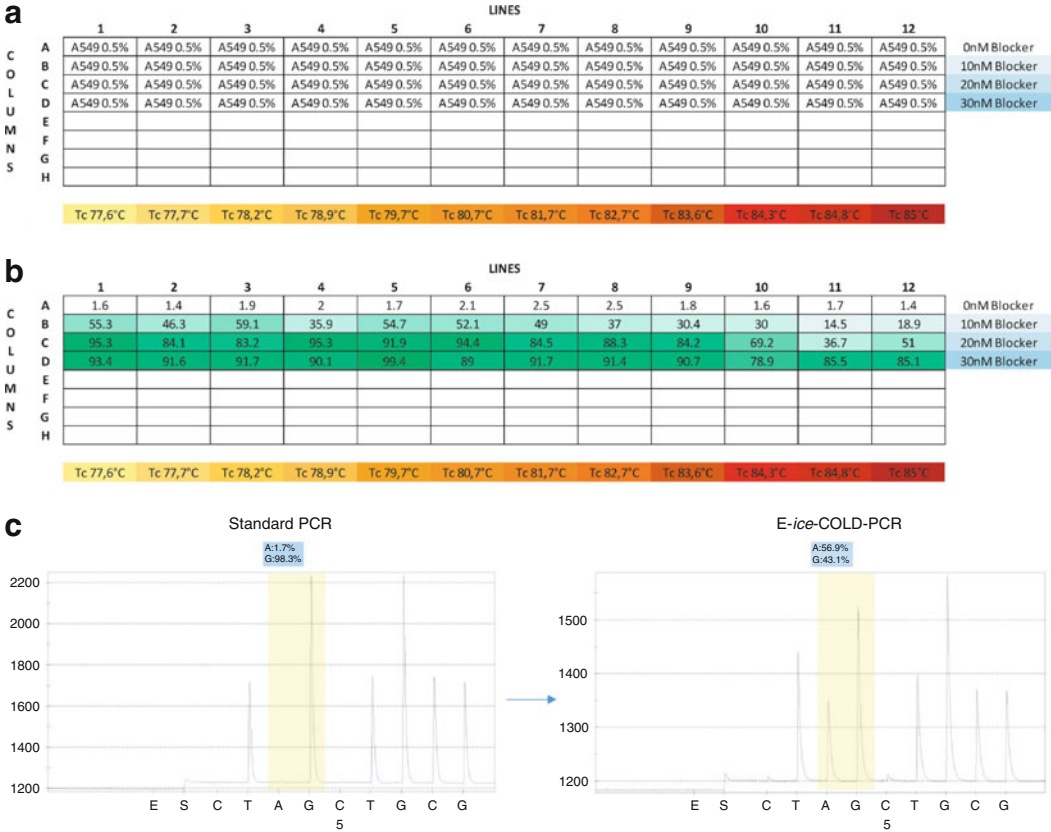


Fig. 3 Example of the determination of the critical temperature for the E-ice-COLD-PCR analyzing the mutation hotspot in *KRAS* codons 12 and 13 and using the 114 bp assay and the 31 bp LNA blocker published in [12]. (a) Example of a plate layout using a *KRAS* c.34G>A mutated cell line (A459) diluted to 0.5 % mutated allele, different blocker concentrations from 0 to 30 nM, and a Tc gradient ranging from 77.6 °C to 85 °C. (b) Quantification of the mutation enrichment obtained after Pyrosequencing for the plate layout presented in (a). The selected Tc is 80 °C. (c) Example of Pyrograms obtained for a 0.5 % mutation fraction of the *KRAS* c.34G>A mutation (A549) after standard and E-ice-COLD-PCR

5. Monitor the effect of the blocker probe on the amplification plots, which are with increasing blocker concentration progressively delayed compared to amplifications without blocker probes. For some Tc and/or concentrations of the blocker probe, the inhibition effect could be very strong resulting in the absence of PCR amplification products and/or in the presence of a large amount of primer dimers.
6. Analyze 10 µL of all PCR amplification products by Pyrosequencing to quantify the enrichment of the mutations (see Subheading 3.4).
7. In the resulting Pyrogram (output format of the Pyrosequencer), the enrichment of the mutation should be

clearly visible for a large range of temperatures and correlate positively with increasing blocker probe concentration. For each blocker probe concentration the optimal T_c should be the temperature where the highest enrichment is achieved together with an excellent quality of the Pyrograms, i.e., single-nucleotide peak intensity higher than 100 units.

8. Test the performance of the E-*ice*-COLD-PCR assay using the same PCR conditions (**steps 3 and 4** above) at the selected T_c using the different blocker probe concentrations (prepared in Subheading 3.3.2, **step 2**) including new concentrations and excluding non-informative concentrations (*see Note 9*). Perform the experiments with the dilutions of the cell line DNA of 1 % and lower dilutions (0.5, 0.1, 0.05, and 0.01 %) (*see step 4* in Subheading 3.2.1) and WT commercial samples in order to determine the limit of detection of the assay and the optimal blocker probe concentration.
9. Perform **steps 5 and 6** on these amplifications.
10. Identify the optimal blocker probe concentration, which corresponds to the concentration where the highest mutation enrichment is achieved for the lowest mutation fraction associated to a high quality of the resulting Pyrogram and where the WT sample shows a mutation value close to 0 % (*see Note 10*).
11. Validate the optimal conditions of the assay using the determined T_c and blocker probe concentration on the whole dilution series of the cell line including also WT and no-template samples.
12. The optimal concentration of the blocker probe depends on the quality and the quantity of DNA present in the analyzed samples and must therefore be adapted to the starting amount of DNA. DNA from FFPE and plasma samples is degraded and this should be taken into account for the choice of LNA blocker probe concentration. Moreover, when the starting amount of DNA decreases, the LNA blocker probe concentration can be decreased simultaneously as the lower number of molecules present in the reaction requires a lower sensitivity. Use the determined concentration of blocker probe for DNA of high quality such as fresh-frozen sample. For DNA from FFPE or circulating DNA, reduce the concentration of the blocker probe by 10–20 nM.
13. Once optimal PCR primer annealing temperature, critical temperature, and concentration of LNA blocker probes have been identified, analyze the different samples of interest by E-*ice*-COLD-PCR assays preferentially in duplicate or triplicate experiments.

3.4 Mutation Detection, Identification, and Quantification by Pyrosequencing

3.4.1 Sample Preparation

This step allows the preparation of the PCR products for Pyrosequencing experiments. It includes the purification and the denaturation of the amplification product, which will be rendered single stranded, and the hybridization of the sequencing primer.

1. Transfer 10 μL of E-ice-COLD-PCR product into a standard skirted 96-well PCR plate and add 40 μL of binding buffer, 2 μL of Sepharose beads, and 28 μL of water. Seal the plate and incubate for 10 min at room temperature under constant mixing (1,400 rpm), which is required for the capture of the biotinylated PCR products by the Sepharose beads.
2. In parallel, prepare the Pyrosequencing plate by diluting 4 pmol of the Pyrosequencing primer into 12 μL of annealing buffer into the respective wells of the PSQ plate. Different Pyrosequencing primers can be used in different wells of the same plate.
3. Fill the four troughs of the vacuum preparation tool with 100 mL of 70 % ethanol, 100 mL 0.2 M NaOH denaturing solution, 120 mL of washing buffer, and 150 mL of water. More washing buffer should be put in the troughs compared to ethanol and NaOH to ensure the complete removal of NaOH, which might inhibit the downstream Pyrosequencing reactions.
4. Turn on the workstation to create a vacuum in the aspiration device (450 mmHg). Immerse the tips in water for several seconds for cleaning. Then, the PCR plate can be removed from the mixer and the binding mix can be aspirated. The beads remain on the filters of the tips.
5. After all binding mix has been aspirated, immerse the tips in ethanol 70 %, NaOH denaturing solution, and washing solution for 5 s, 5 s, and 10 s, respectively. Let the vacuum dry the beads for 10 s and turn it off above the PSQ plate.
6. Next, the filter tips can be immersed in the annealing mix of the PSQ plate in order to release the beads into the wells.
7. Heat the sequencing plate for 2 min at 80 °C on a thermoplate placed on a heating device for single DNA strand denaturation. Cool the plate at room temperature for 5 min to allow annealing of the sequencing primer.

3.4.2 Pyrosequencing Reaction

1. Create a new Pyrosequencing run on the Pyrosequencer using the PyroMark MD software. Indicate the genotyping assay used in the selected wells and the name and/or conditions of each wells.
2. Fill the tips of the cartridge with the corresponding reagents and enzyme mix and dNTPs avoiding bubble formation. 180 μL of reagents and enzyme mix are sufficient for a full

plate and at least 150 μL of each dNTPs should be used. dNTPs are not limited in the kit and can be used in excess (*see* **Note 11**).

3. Deposit a sealed Pyrosequencing test plate and the reagents' cartridge in the Pyrosequencer and perform the dispensation test to verify if the dispensing tips are working properly. Droplets should be clearly visible and homogeneous and the tips should be changed if necessary.
4. Remove the test plate, put the cooled prepared Pyrosequencing plate, and start the run. The length of a run is proportional to the number of dispensations (1/min) and a genotyping run should not exceed 30 min.
5. After the end of the sequencing run, rerun a dispensation test to verify if some tips became blocked during the run. If the tips were blocked or if no signal was present because of exhausted or degraded reagents, the Pyrosequencing plate should be resuspended in 20 μL of binding buffer, and re-purified by the protocol described in **step 2** in Subheading 3.4.1.
6. Analyze the results with the PyroMark MD software in AQ mode, which allows accurate quantification of the mutation(s) of interest. For more complicated patterns and mutation hotspot regions, the Pyrograms need to be manually analyzed for identification of the mutation and the intensity peaks have to be manually processed or with a help of a "homemade" program for mutation quantification (*see* Subheading 3.2.4).
7. Export the results in .txt format to enable further treatment and analysis with statistical or graphical software such as Excel®.
8. Remove the cartridge and clean the tips with distilled water. Remaining reagents can be preserved at 4 °C and reused while dNTPs can be discarded.

An example of Pyrograms before and after enrichment of *KRAS* mutation in colorectal cancer samples is shown in Fig. 3c.

4 Notes

1. The best performing assay designed for the E-*ice*-COLD-PCR application in the region of interest should be used for the quantitative PCR omitting of course the blocker probes used for the selective amplification of mutated alleles. The quantity of amplifiable DNA of the region of interest can be accurately determined using a dilution series of commercial genomic DNA ranging from 25 ng to 25 μg as calibration standard. Typical quantitative PCR reaction conditions are 1 $\times$ HotStar Taq buffer supplemented with 1.6 mM MgCl_2 , 200 mM of

each dNTP, 1.0 U of HotStar Taq polymerase, 2 μ M of SYTO9, and 200 nM of forward and reverse primers in a total volume of 25 μ L. Typical quantitative PCR cycling conditions include an initial denaturation step performed 10 min at 95 °C, followed by 55 cycles of 30-s denaturation at 95 °C, 20 s at annealing temperature, and 10-s elongation at 72 °C. Fluorescence acquisition is performed during each cycle at the end of the primer annealing step. Final steps include a melting curve performed from 65 °C to 97 °C and cooling to 40 °C.

2. If no cell lines with a mutation in the region of interest can be identified, a validated sample presenting a mutation in the region of interest could be used instead. Thus, whole-genome amplification (WGA) should be performed on the sample using REPLI-g Midi kit (Qiagen) according to the manufacturer's instructions and mixed with a WT WG amplified sample to obtain the different mutation fraction dilution. It is not recommended to mix a WGA mutant sample with an unamplified WT sample, as due to the different DNA fragment sizes, differences in the amplification efficiency might occur.
3. Quantification of a DNA sample based on the measurement of the absorption/optical density only is not sufficiently accurate and must be complemented by more precise quantification methods such as intercalating dyes or quantitative real-time PCR.
4. The dilutions series allows the determination of the limit of detection of each E-ice-COLD-PCR assay.
5. Some regions can be extremely homologous to the region of interest such as pseudogenes, for example the *KRAS* pseudogene 1 (*KRASPI*), and it is primordial to have at least five nucleotides specific only to the region of interest on the 3' end of at least one primer.
6. Microsoft Excel Visual Basic Applications (VBA) are a good platform for the development of an automated treatment of raw peak intensities for mutation detection, identification, and quantification. An example of a developed MS VBA for the analysis of the *KRAS* codon 12 and 13 mutation hotspot is shown in [12].
7. The melting curve analysis can also be performed as high-resolution melting (HRM) analysis, which can also be used for mutation detection at the end of the E-ice-COLD-PCR prior to downstream experiments. At least 20 acquisitions per degree should be included in the melting program to achieve a satisfying analytic resolution.
8. The use of a gradient real-time thermocycler such as LightCycler® 96 Real-Time PCR system can significantly improve the setup efficiency of E-ice-COLD-PCR assays as amplification plots and primer dimer formation can be monitored

directly during the PCR, before Pyrosequencing experiments. If no gradient real-time thermocycler is available, a standard gradient thermocycler can be used for the optimization steps.

9. If the first experiments for the determination of the T_c show no mutation enrichment for the lowest tested blocker concentration, these concentrations should be excluded from further analysis while if the quality of the Pyrogram remains high for the highest tested concentration, higher concentrations could and should be tested in the following experiments.
10. The limit of detection corresponds to the lowest mutation fraction where mutation enrichment is still visible compared to the WT sample. As Pyrosequencing presents a resolution of 5–10 %, a mutation difference of at least 10 % between WT and mutated samples after E-*ice*-COLD-PCR should be taken as cutoff to maximize the sensitivity and the specificity of the E-*ice*-COLD-PCR assay. After E-*ice*-COLD-PCR, a WT sample can present with a higher mutation background compared to a standard PCR. This is due to the stringency/aggressiveness of the method and the blocker probe concentration should be optimized to minimize this effect.
11. Reaction volumes and required input of the amplification product might vary depending on the model of the Pyrosequencing machine.

Acknowledgments

We thank Florence Mauger (CNG), Helene Myrtue Nielsen (Aarhus University), and Steven McGinn (CNG) for critical reading and improvement of the manuscript.

References

1. Wittwer CT (2009) High-resolution DNA melting analysis: advancements and limitations. *Hum Mutat* 30:857–859
2. Ugozzoli L, Wallace RB (1991) Allele-specific polymerase chain reaction. *Methods* 2:42–48
3. Ogino S, Kawasaki T, Brahmandam M et al (2005) Sensitive sequencing method for KRAS mutation detection by Pyrosequencing. *J Mol Diagn* 7:413–421
4. Ahmadian A, Ehn M, Hober S (2006) Pyrosequencing: history, biochemistry and future. *Clin Chim Acta* 363:83–94
5. Ronaghi M, Shokralla S, Gharizadeh B (2007) Pyrosequencing for discovery and analysis of DNA sequence variations. *Pharmacogenomics* 8:1437–1441
6. Ronaghi M, Uhlen M, Nyren P (1998) A sequencing method based on real-time pyrophosphate. *Science* 281:363–365
7. Ronaghi M (2001) Pyrosequencing sheds light on DNA sequencing. *Genome Res* 11:3–11
8. Milbury CA, Li J, Makrigiorgos GM (2009) PCR-based methods for the enrichment of minority alleles and mutations. *Clin Chem* 55:632–640
9. Li J, Wang L, Mamon H et al (2008) Replacing PCR with COLD-PCR enriches variant DNA sequences and redefines the sensitivity of genetic testing. *Nat Med* 14:579–584
10. Milbury CA, Li J, Makrigiorgos GM (2009) COLD-PCR-enhanced high-resolution melting

- enables rapid and selective identification of low-level unknown mutations. *Clin Chem* 55:2130–2143
11. Milbury CA, Li J, Makrigiorgos GM (2011) Ice-COLD-PCR enables rapid amplification and robust enrichment for low-abundance unknown DNA mutations. *Nucleic Acids Res* 39:e2
 12. How Kit A, Mazaleyrat N, Daunay A et al (2013) Sensitive detection of KRAS mutations using enhanced-ice-COLD-PCR mutation enrichment and direct sequence identification. *Hum Mutat* 34:1568–1580
 13. Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23:1289–1291
 14. Untergasser A, Cutcutache I, Koressaar T et al (2012) Primer3—new capabilities and interfaces. *Nucleic Acids Res* 40:e115
 15. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31:3406–3415
 16. Forbes SA, Bindal N, Bamford S et al (2011) COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res* 39:D945–D950
 17. Dunker J, Larsson U, Petersson D et al (2003) Parallel DNA template preparation using a vacuum filtration sample transfer device. *Biotechniques* 34:862–868
 18. Zhou GH, Gotou M, Kajiyama T et al (2005) Multiplex SNP typing by bioluminometric assay coupled with terminator incorporation (BATT). *Nucleic Acids Res* 33:e133
 19. You Y, Moreira BG, Behlke MA et al (2006) Design of LNA probes that improve mismatch discrimination. *Nucleic Acids Res* 34:e60

Free ebooks ==> www.Ebook777.com

Chapter 8

Analysis of Mutational Hotspots in Routinely Processed Bone Marrow Trephines by Pyrosequencing®

Stephan Bartels and Ulrich Lehmann

Abstract

Formalin-fixed, paraffin-embedded (FFPE) bone marrow trephines are widely used in pathology, because they best preserve the morphological details of the bone marrow. However, DNA isolated from FFPE material is fragmented, limiting the size of amplification products, which is a challenge for all sequencing applications.

Pyrosequencing® is a quantitative and sensitive method for the detection of single-nucleotide variations (SNVs) in DNA samples. Pyrosequencing can easily be performed in a 96-well-plate format with a cost-effective medium-sized throughput.

This chapter provides a general outline of SNV detection in FFPE bone marrow trephines, including a detailed protocol of the Pyrosequencing procedure and guidelines for the design of new assays and evaluation of Pyrograms. The strengths of this approach are discussed using myeloproliferative neoplasms as an example.

Key words Pyrosequencing®, Bone marrow trephines, Mutational hotspots, MPN, Assay design

1 Introduction

Myeloproliferative neoplasms (MPNs) are divided into nine subcategories in the 2008 WHO classification [1]. The four most important of them, chronic myelogenous leukemia (CML), polycythemia vera (PV), essential thrombocytosis (ET), and primary myelofibrosis (PMF), are chronic, clonal hematopoietic stem cell disorders, characterized by the presence of an excess of terminally differentiated cells [2]. ET, PV, and PMF are also called “BCR-ABL1-negative MPN,” to distinguish them from chronic myeloid leukemia [3]. These diseases have a specific risk for a progress of bone marrow fibrosis and leukemic transformation into a secondary acute myeloid leukemia (2nd AML) [3, 4].

Especially the discrimination between ET and the prefibrotic phase of PMF is important, because PMF patients have a higher risk of progression to myelofibrosis or leukemic transformation [5].

However, the distinction based only on bone marrow morphology and clinical data is not always possible [5, 6]. Therefore, it is important to combine the histological classification with sensitive molecular analysis methods for diagnosis, prognosis, and treatment decisions.

Archival bone marrow trephines are the gold standard for the diagnosis of many bone marrow-derived hematological diseases. They provide morphological details and enable immunohistochemical characterization of all cell types of the bone marrow, and their long-term storage is inexpensive [7]. Usually trephines are formalin fixed, decalcified, and paraffin embedded (FFPE). In contrast to bone marrow aspirates FFPE specimens suffer from fragmentation of nucleic acids due to the fixation process [8]. DNA isolated from FFPE samples often contains only fragments of a few hundred base pairs. This might be limiting for Sanger sequencing of longer fragments, but does not interfere with Pyrosequencing. Additionally, Pyrosequencing has a superior limit of detection compared with Sanger sequencing, which is a great advantage analyzing heterogeneous tumor samples [9].

A milestone in the molecular description of MPNs was the discovery of the *JAK2* V617F gain-of-function mutation in 2005 [10]. *JAK2* is frequently mutated in MPNs, nearly all PV cases are V617F mutated (~96 %, V617F-negative PV patients commonly carry other *JAK2* mutations, e.g., exon 12 deletions), and approximately 50–60 % of patients with ET or PMF [3, 11].

Recent genome-wide analysis of MPN patients revealed pathogenic mutations in different classes of genes with different cellular functions. Mutated genes are involved in cell signaling pathways (*JAK2*, *MPL*, *FLT3*), epigenetic regulation (*IDH1/2*, *DNMT3A*, *ASXL1*), and RNA maturation (*SRSF2*, *U2AF1*, *SF3B1*) [12, 13].

Many of the genes, which are frequently mutated in MPNs, exhibit mutational hotspots, which are suitable for SNP analysis by Pyrosequencing assays. For example *SRSF2* mainly shows missense mutations, insertions, and deletions involving Proline 95 [14]. *IDH1* missense mutations affect Arginine 132 and *IDH2* missense mutations affect Arginine 140 and 172 [15, 16]. Molecular analysis of MPN-related genes complements the morphological based examination of bone marrow trephines. The obtained results can support or specify the diagnosis and give useful further information concerning prognosis and treatment decisions.

SRSF2 for instance is frequently mutated in PMF but not in ET. Thus *SRSF2* mutations can discriminate early-phase PMF from ET [17]. Additionally *SRSF2* mutations are a negative prognostic marker for shortened overall survival and leukemia-free survival [14].

2 Materials

2.1 Tissue Processing

1. Bone marrow fixative: 64 % methanol (v/v), 4.48 M formaldehyde, 1.6 mM sodium hydrogen phosphate pH 7.4, 7.4 mM glucose.
2. Decalcification solution: 270 mM Tris-HCl, 270 mM EDTA, pH 7.4.
3. 60 mM sodium hydrogen phosphate, pH 7.0–7.2.
4. Paraffin.

2.2 DNA Isolation

1. Microtome.
2. AppliClear, a xylene substitute (AppliChem).
3. Proteinase K buffer: 50 mM Tris-HCl pH 8.1, 1 mM EDTA, 0.5 % Tween 20.
4. Proteinase K (20 mg/mL).
5. Aerosol filter tips (*see* **Note 1**).
6. Thermoshaker.
7. Phenol/chloroform.
8. Refrigerated tabletop centrifuge.
9. Chloroform.
10. RNase A (10 mg/mL).
11. Sodium acetate (3 M).
12. Dextran (20 % solution).
13. Ethanol.
14. TE buffer: 10 mM Tris-HCl, pH 8.1, 1 mM EDTA.

2.3 Pyro-PCR

1. Taq-Polymerase with buffer and $MgCl_2$ (e.g., “Platinum-Taq,” Life Technologies).
2. Primers (*see* **Note 3**): *SRSF2* primers:
5'-GAG CTG CGG GTG CAA ATG-3' and
5'-CGT ACC TGC GGG GTG GC-3'.
3. dNTPs.
4. Aerosol filter tips.
5. PCR tubes or plates.
6. PCR machine (e.g., Personal Cycler, Biometra).

2.4 Pyrosequencing

1. Pyrosequencing system (PyroMark[®] MD, Qiagen, Hilden, Germany).
2. Pyrosequencing reagent kit (Qiagen).
3. Reagent dispensing tips (Qiagen).

4. Capillary dispensing tips (Qiagen).
5. Q-CpG software (Qiagen).
6. Primers for Pyrosequencing: *SRSF2* primer:
F: 5'-GCG CTA CGG CCG C-3' and
R: 5'-GTG GTG TGA GTC CGG-3'.
7. PCR plates.
8. Pyrosequencing plates (Qiagen).
9. Streptavidin Sepharose HP beads (GE Healthcare, Uppsala, Sweden).
10. Vacuum workstation with the appropriate filter tips and troughs (Qiagen).
11. Binding buffer: 10 mM Tris-HCl, pH 7.6, 2 M NaCl, 1 mM EDTA, 0.1 % v/v Tween 20.
12. Denaturation solution: 0.2 N NaOH.
13. Washing buffer: 10 mM Tris-acetate, pH 7.6.
14. Annealing buffer: 20 mM Tris-acetate, pH 7.6, 2 mM Mg acetate.
15. Microplate mixer (max. 2,000 rpm).
16. Heating plate or thermoblock for the denaturation step.
17. Thermoplate for sample preparation (Qiagen).

3 Methods

3.1 Fixation and Decalcification of Bone Marrow Trephines

The optimal fixation procedure uses buffered formalin for 24 h at 4 °C in the dark. This cannot always be achieved under routine conditions but some deviation from these optimal conditions (longer duration, elevated temperature compared to room temperature) seems to be tolerable. The volume of the fixative should be much greater (ten times or more) than the volume of the specimen. The pH of the EDTA solution is adjusted to pH 7.4 using 60 mM sodium hydrogen phosphate pH 7.0–7.2. The solution is stirred constantly and changed every day. The decalcification in the neutral EDTA buffer can take several days. But conservation of nucleic acids is guaranteed under these conditions. By using an ultrasonic bath this time can be considerably shortened [18]. After fixation and decalcification the tissue is embedded in liquid paraffin (62 °C), which solidifies by cooling down to room temperature.

3.2 DNA Isolation

FFPE biopsies are cut using a conventional microtome. Depending on the biopsy and the requested yield of DNA one till six slides of 15 µm are cut.

1. Lyse the samples overnight with 1,000 µL of Proteinase K buffer and 50 µL Proteinase K in a Thermoshaker at 55 °C with 800 rpm.

2. Transfer the lysate into a new tube by pricking the overlying paraffin cap with an aerosol filter tip.
3. Add 750 μL phenol/chloroform to the lysate for extraction, vortex thoroughly, and centrifuge at $12,000\times g$ at 4°C for 15 min.
4. Transfer the aqueous supernatant into a new tube and repeat the extraction until the white interphase is completely removed.
5. Add 750 μL chloroform, vortex thoroughly, and centrifuge at $12,000\times g$ at 4°C for 3 min.
6. Transfer the aqueous supernatant into a new tube and add 5 μL RNase A, vortex thoroughly, and incubate in a Thermoshaker at 37°C for 30–60 min.
7. Perform additional phenol/chloroform extraction (1 $\times$).
8. Add 750 μL chloroform, vortex thoroughly, and centrifuge at $12,000\times g$ at 4°C for 3 min.
9. Transfer the aqueous supernatant (approx. 800 μL) into a new tube and add 1/10 Vol. of sodium acetate and 1,000 μL of 100 % ethanol. Precipitate the DNA overnight at -20°C .
10. Centrifuge at $12,000\times g$ at 4°C for 20 min.
11. Discard the supernatant and wash the pellet with 500 μL of 70 % ethanol.
12. Centrifuge at $12,000\times g$ at 4°C for 5 min.
13. Discard the supernatant, air-dry the pellet for 5 min, and elute in 50–100 μL TE buffer.

Organic extraction of DNA requires more practical skills and is quite time consuming, but the procedure yields high amounts of pure DNA without the risk of “overloading” a column with sample material and subsequent clogging. Especially in case of fragmented DNA it is an advantage, because often several sections must be extracted to get enough DNA for robust and reliable amplification.

Also numerous DNA isolation kits are commercially available. We use the “DNeasy Blood and Tissue Kit” (Qiagen) for diagnostic applications, which performs very well.

3.3 Assay Design

We strongly recommend designing two complementary assays for mutational hotspots, a forward and a reverse assay. Especially tumor samples often have low frequencies of mutated strands and percentage of tumor cells in a sample is often low or unclear. Two independent sequencing reactions lower the risk of false-positive or false-negative results and prevent wrong diagnostic and therapeutic decisions. Furthermore two independent assays are helpful for analyzing difficult Pyrograms (e.g., from sequences, which contain homopolymers).

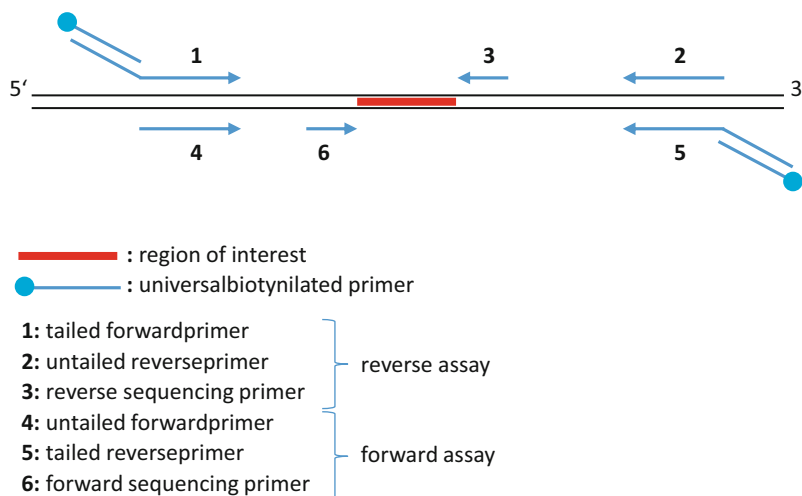


Fig. 1 Schematic view of a forward and a reverse assay for mutation detection with Pyrosequencing. Every assay includes a Pyro-PCR primer pair (one primer is tailed in this schema, *see* text for details) and a sequencing primer

Primer pairs for Pyro-PCR (*see* Subheading 3.4) can be designed including a biotin-tag gene-specific primer. Another possibility is to use a universal biotinylated primer (*see* Fig. 1).

The sequence of this primer does not exist in the human genome. With the complement tail on one of the Pyro-PCR primers, the universal biotinylated primer can be used for every assay. Biotinylated primers are expensive, so the advantage is that especially for the test of the primers (*see* **Note 3**), no direct biotinylated primers need to be ordered. We use directly biotinylated Pyro-PCR primers only in validated diagnostic assays, which are frequently performed (*see* **Note 2**).

Primers can be designed by the “Assay Design Software” (Biotage, now Qiagen). However, in our experience, the manual selection of primers supported by a conventional primer design program (to calculate T_{Ann} and GC content and identify potential primer dimers and hairpin structures) is equally efficient and sometimes even better. We routinely use Primer Express (formerly Applied Biosystems, now Thermo Fisher Scientific) or Primer 3 (<http://frodo.wi.mit.edu/>) for this purpose. The following negative controls should be included during assay setup to detect undesired self-priming reactions that may create strong background signals:

1. Inclusion of only the sequencing primer in the Pyrosequencing reaction.
2. Inclusion of only the biotinylated primer in the Pyrosequencing reaction.

3. Inclusion of only the sequencing primer and the biotinylated primer in the Pyrosequencing reaction.
4. Inclusion of only the PCR product.

3.4 Pyro-PCR

1. 10–50 ng of FFPE-DNA are amplified in a total volume of 25 μL using 0.5 units Platinum-Taq, 10 pmol of each SRSF2 primer (*see* above Subheading 2.3), and 200 μM dNTPs.
2. The optimal MgCl_2 concentration and annealing temperature have to be determined empirically (*see* **Note 3**).
3. Perform 45 cycles to reach the plateau and obtain enough PCR product for sequencing (especially in case of fragmented DNA samples), furthermore to exhaust as much primers as possible. This will reduce the background signal of the sequencing run.
4. After completion of the PCR step, 4 μL of the reaction mixture is resolved on a polyacrylamide gel and stained with ethidium bromide in order to verify the identity and purity of the PCR product.

3.5 Pyrosequencing

1. Fill the troughs at the vacuum preparation station. Trough 1: 70 % v/v ethanol; trough 2: 0.2 M NaOH; trough 3: 1 $\times$ washing buffer; and trough 4: distilled water. Switch on heating block (set to 80 $^{\circ}\text{C}$) and place thermoplate onto the heating block.
2. Dispense into a 96-well plate the following mixture for every sample:
 - (a) 47 μL binding buffer.
 - (b) 3 μL streptavidin-Sepharose beads (mix well before use).
 - (c) Up to 10 μL PCR product (usually 5 μL (*see* Subheading 3.6)).
 - (d) Add HPLC water to a total volume of 80 μL .
3. Shake the plate with 1,400 rpm on a microplate mixer for 5 min at room temperature.
4. Meanwhile, prepare the Pyrosequencing plate by adding 11.5 μL annealing buffer for each reaction well and 0.5 μL (5 pmol) Pyrosequencing primer.
5. The order of samples within the wells of the PCR plate must be identical to the order in the binding reaction and the Pyrosequencing plate.
6. Turn on the pump of the vacuum preparation station, close the valve to create a vacuum, and wash the tips for 10 s in trough 4 (distilled water). Next, check the filter tips of the aspiration device by aspirating water from a separate PCR plate containing 100 μL HPLC water in each well. The water should

be aspirated from every well within a few seconds. If this is not the case, all corresponding tips must be replaced.

7. Aspirate the Sepharose beads from the sample plate for at least 10 s and carefully remove the aspiration device. Take care not to touch the edge of the PCR plate wells. This will result in a loss of Sepharose beads.
8. Immerse the aspiration device for 10 s in trough 1 (*see above; step 1*), and then place the device in trough 2 for 20 s and trough 3 for 30 s. Avoid touching the edges of the troughs.
9. Hold the aspiration device above the Pyrosequencing plate, turn off the vacuum, and check that the air pressure has reached a normal level. Take care to ensure that the display has reached zero before proceeding! Immerse the tips in the annealing mix afterwards. Shake or tap gently for 15–30 s to release the beads into the wells. Make sure that all beads are shaken off by holding the plate against a light source. If the aspiration device is lowered too fast, the annealing mixture might be drawn off.
10. Place the Pyrosequencing plate on top of the heating block (80 °C), cover with the thermoplate, and incubate for 2 min. Allow to cool for 5–10 min.
11. In the meantime, start the “PyroMark™ MD” software system selecting the appropriate assay setup file. Register all samples in the 96-well plate in the preset worksheet. When selecting “View/Run,” the program calculates the required amount of enzyme, substrate, and nucleotides (depending on the number of samples and the read length for every individual sample).
12. Dispense an appropriate amount of enzyme, substrate, and nucleotide into the cartridges. Avoid the ingress of air bubbles, as this may cause an obstruction in the fine dispensation needles and will result in dispensation errors in the Pyrogram. Qiagen recommends centrifugation of the nucleotides before use for 5 min at full speed at room temperature, to avoid the transfer of any particles into any of the six cartridges.
13. Place the holder with the filled cartridges and the Pyrosequencing plate into the analyzer, close the reaction chamber via the software, close the lid of the analyzer manually, and start the run via the software (*see Note 4*).
14. After finishing the run, prompt and careful cleaning of the cartridges is of utmost importance. To avoid clogging of the needles fill the cartridges with HPLC water and push with a finger on the upper orifice to create a water jet.

3.6 Guidelines for Pyrogram Evaluation

1. If the substrate is dispensed at the very beginning, it should not generate a signal (or only a peak much smaller than the nucleotide signal). A large substrate peak indicates pyrophosphate contamination, and should be avoided.

2. Peaks should be sharp and slender, and should reach the baseline after less than 30 s (halfway before the next dispensation). A wide peak indicates too much PCR product input in the sequencing reaction.
3. Single-nucleotide peaks should be very similar for C, G, and T and slightly larger for A.
4. No additional peaks should appear between the regularly spaced nucleotide peaks.
5. The average peak height should be clearly above background, at least 25 relative light units (RFUs). 50 units above the background are clearly much better and achievable with the PyroMark MD system we use. Values differ for the other two systems from Qiagen (PyroMark Q24 and Q96 ID, *see* Chapter 2).
6. The measured peaks should fit well into the calculated histogram.

The “PyroMark™ MD” software will generate example Pyrograms, when the sequence to analyze is entered in the software. This virtual Pyrograms can be compared with the generated Pyrograms. Another possibility is to create virtual Pyrograms with free software tools, for example pyromaker (<http://pyromaker.pathology.jhmi.edu/index.php>, *see* Chapter 3). Helpful is also to establish plasmid controls from wild-type and mutated DNA samples. Thereby also different frequencies of mutated strands can be determined (*see* Chapter 4). The Pyrogram may look completely different, depending on the type and frequency of the mutation. For an example of Pyrogram evaluation *see* Fig. 2.

4 Notes

1. Repeated amplification of the same target sequence from limited amounts of starting material poses a serious risk of cross-contamination of samples. For this reason, strict guidelines concerning the handling of samples before and after amplification and the cleaning of all instruments must be enforced and carried out by all personnel involved. We perform all pre-amplification steps in a separate laboratory (“pre-PCR area”). Everything used in this laboratory (including lab coats, pens, and notepads) are dedicated exclusively to this room and are strictly separated from the post-PCR area. Plastic lab-ware and benches are cleaned regularly using a 3 % v/v hypochlorite solution. PCR products are subsequently analyzed in a separate laboratory (“post-PCR area”). Under no circumstances should amplified samples or equipment from this working area be brought back to the pre-PCR area. Aerosol filter tips are also an absolute necessity in the pre-PCR area.

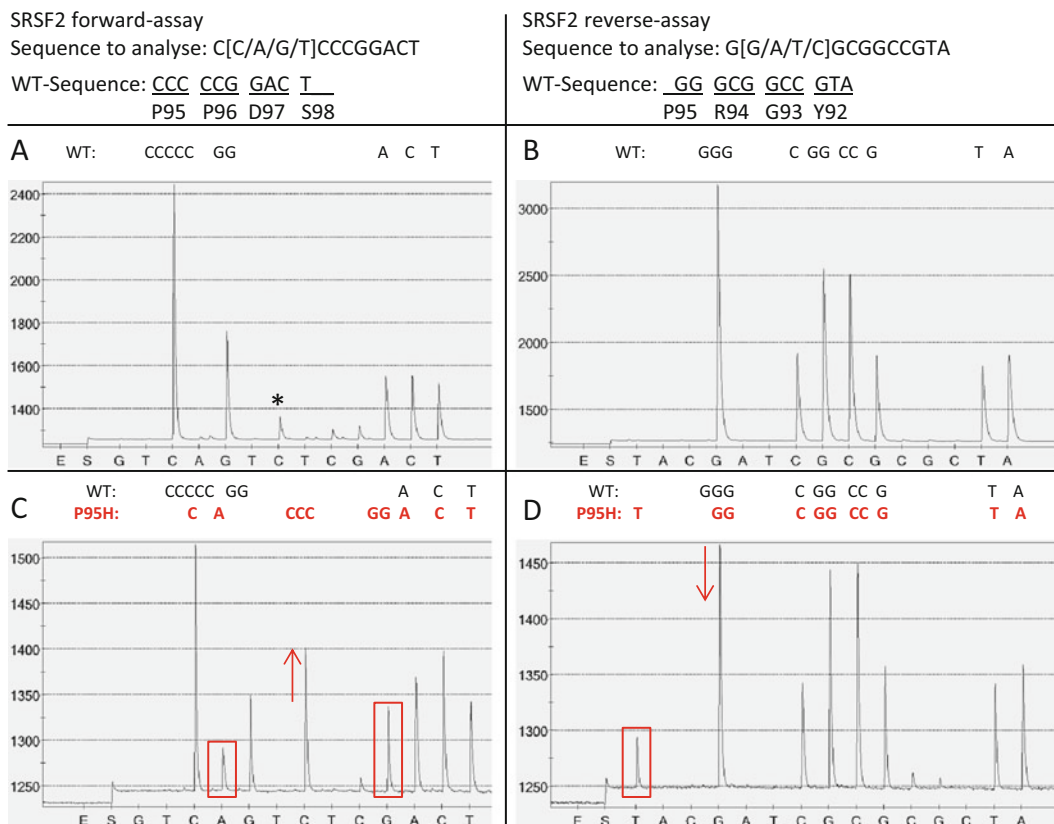


Fig. 2 The primary output of the “PyroMark™ MD” software. In (a) and (b) patient DNA was sequenced for *SRSF2* mutational hotspot (Proline 95) and found to be wild type. (c) and (d) show a patient, which has a P95H mutation in the *SRSF2* gene, the most common mutation in this gene [12]. The differences in the Pyrograms are marked with red boxes and arrows. (a) and (c) show the forward sequencing assays, (c) and (d) the reverse sequencing assay. Above the Pyrograms the sequence to analyze from the specific assay and the wild-type sequence are indicated. The asterisk shows a peak from a base, where the specific nucleotide was not completely incorporated in all strands at the former injection (C in this case). This is a problem when the sequence contains homopolymers (5 C's in this case). This leads to a decline of the homopolymer peak. Next time the same nucleotide is injected, the polymerase will incorporate the nucleotide in the remaining strands and generates a new small peak

2. Biotinylated primers are more labile than normal primers. Repeated thawing and freezing should be avoided by preparation of aliquots containing the stock solution. High purity (HPLC grade) is also required since unbound biotin competes for binding to the streptavidin-agarose beads and, thereby, reduces the efficiency of the purification step. For most assays, we use a universal biotinylated primer and a complementary tag at one of the two PCR primers [19]. This saves quite a lot of money as it avoids the ordering of dozens of biotinylated primers, which are sometimes not used up within

the shelf life of the reagent. Under most circumstances, the tag and the universal primer do not interfere with the amplification of the sequence of interest. Only for frequently used assays and if the tag interferes with the specificity and/or efficiency of the PCR we order assay-specific biotinylated primers.

3. Primer optimization, testing different annealing temperatures and MgCl_2 concentrations, is strongly recommended. We routinely test three to six different temperatures around the calculated optimal T_{Ann} at two different MgCl_2 (1.5 and 2.5 mM) concentrations and select the reaction parameters giving the strongest band and no or only spurious side products.
4. Before each run, perform a dispensation test: Place a Pyrosequencing plate sealed with a plastic foil inside the reaction chamber and start a dispensation test using the software. Six discrete droplets should be clearly visible and each should be positioned at the center of a well. This ensures that the dispensation is working at the beginning of the analysis. The test does not, however, prevent occasional obstruction of one or several dispensation needles during the Pyrosequencing run, but it does reduce the risk of a complete failure of the sequencing run.

References

1. Tefferi A, Vardiman J (2008) Classification and diagnosis of myeloproliferative neoplasms: the 2008 World Health Organization criteria and point-of-care diagnostic algorithms. *Leukemia* 22:14–22
2. Cleary C, Kralovics R (2013) Molecular basis and clonal evolution of myeloproliferative neoplasms. *Clin Chem Lab Med* 51:1889–1896
3. Tefferi A (2010) Novel mutations and their functional and clinical relevance in myeloproliferative neoplasms: JAK2, MPL, TET2, ASXL1, CBL, IDH and IKZF1. *Leukemia* 24:1128–1138
4. Zhang S-J, Rampal R, Manshouri T et al (2012) Genetic analysis of patients with leukemic transformation of myeloproliferative neoplasms shows recurrent SRSF2 mutations that are associated with adverse outcome. *Blood* 119:4480–4485
5. Tefferi A (2013) Primary myelofibrosis: 2013 update on diagnosis, risk-stratification, and management. *Am J Hematol* 88:141–150
6. Buhr T, Hebeda K, Kaloutsis V et al (2012) European Bone Marrow Working Group trial on reproducibility of World Health Organization criteria to discriminate essential thrombocythemia from prefibrotic primary myelofibrosis. *Haematologica* 97:360–365
7. Hasemeier B, Geffers R, Bartels S et al (2013) Archival bone marrow trephines are suitable for high-throughput mutation analysis using next generation sequencing technology. *Haematologica* 98:e115–e116
8. Dietrich D, Uhl B, Sailer V et al (2013) Improved PCR performance using template DNA from formalin-fixed and paraffin-embedded tissues by overcoming PCR inhibition. *PLoS One* 8:e77771
9. Harrington CT, Lin EI, Olson MT et al (2013) Fundamentals of pyrosequencing. *Arch Pathol Lab Med* 137:1296–1303
10. James C, Ugo V, Le Couédic JP et al (2005) A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. *Nature* 434:1144–1148
11. Soriano G, Heaney M (2013) Polycythemia vera and essential thrombocythemia: new developments in biology with therapeutic implications. *Curr Opin Hematol* 20:169–175
12. Yoshida K, Sanada M, Shiraishi Y et al (2011) Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* 478:64–69
13. Murati A, Brecqueville M, Devillier R et al (2012) Myeloid malignancies: mutations, models and management. *BMC Cancer* 12:304

14. Lasho TL, Jimma T, Finke CM et al (2012) SRSF2 mutations in primary myelofibrosis: significant clustering with IDH mutations and independent association with inferior overall and leukemia-free survival. *Blood* 120:4168–4171
15. Brecqueville M, Rey J, Bertucci F et al (2012) Mutation analysis of ASXL1, CBL, DNMT3A, IDH1, IDH2, JAK2, MPL, NF1, SF3B1, SUZ12, and TET2 in myeloproliferative neoplasms. *Genes Chromosomes Cancer* 51:743–755
16. Shih AH, Chung SS, Dolezal EK et al (2013) Mutational analysis of therapy-related myelodysplastic syndromes and acute myelogenous leukemia. *Haematologica* 98:908–912
17. Lehmann U, Bartels S, Hasemeier B et al (2013) SRSF2 mutation is present in the hypercellular and prefibrotic stage of primary myelofibrosis. *Blood* 121:4011–4012
18. Reineke T, Jenni B, Abdou MT et al (2006) Ultrasonic decalcification offers new perspectives for rapid FISH, DNA, and RT-PCR analysis in bone marrow trephines. *Am J Surg Pathol* 30:892–896
19. Issa JP, Gharibyan V, Cortes J et al (2005) Phase II study of low-dose decitabine in patients with chronic myelogenous leukemia resistant to imatinib mesylate. *J Clin Oncol* 23:3948–3956

Chapter 9

Analysis of Copy Number Variation by Pyrosequencing® Using Paralogous Sequences

Marianne Kristiansen Kringen

Abstract

The determination of gene copy numbers of highly similar genes is difficult with conventional PCR-based methods. However, by amplification of similar genes in the same PCR reaction followed by Pyrosequencing®, one may distinguish the genes based on a single-nucleotide difference. The ratio between the peak heights of gene-specific nucleotides obtained in the Pyrosequencing reaction may thereby be used to calculate the relative copy numbers of target genes. This method is easy and cost effective compared to other methods, and allows for the determination of copy numbers of genes that were previously difficult to achieve.

Key words Copy number variation, PCR, Pyrosequencing®, Paralogous sequences, *ABCC6*

1 Introduction

Copy number variation (CNV) refers to a structural variation in the genome that results in the loss (deletion) or gain (duplication) of genome sequences. These segments may range from 1 kb to 5 Mb in length. CNVs are rather common in the human genome, corresponding to about 12 % of gene variation in the genome, and may cause gene dosage differences between individuals [1]. There are several methods for determination of gene CNV, the most common being real-time quantitative PCR of a target gene relative to an endogenous reference gene that is known to be present in two copies in the genome [2]. This method requires a specific amplification of an ~60 bp fragment that is unique in the genome. However, the estimation of CNVs of highly similar genes (e.g., pseudogenes and parent genes) is generally difficult and cannot be done with this method. Pyrosequencing® has been shown to be a helpful tool to differentiate between highly similar genes. The principle is to design PCR primers that amplify two or more DNA sequences in the same PCR reaction. Similarly, a Pyrosequencing primer is designed to bind to all PCR fragments. Thereby, in the following Pyrosequencing reaction, by adding one nucleotide at a

time in a preferential order, one may distinguish the fragments from each other by only one nucleotide difference. Their relative abundance (copy numbers) can be calculated from the ratios between peak heights obtained from the Pyrogram®. For example, if there are equal copies of two genes, the ratios between their specific nucleotide peak heights will equal one.

An example of a gene that is commonly deleted or duplicated in the genome is the cytochrome P450 2D6 (*CYP2D6*) gene, in which the gene product is responsible for the degradation of a wide range of drugs [3]. CNV of the *CYP2D6* gene has a great impact on drug response for the individual patient and is therefore routinely investigated. The *CYP2D6* gene is located on chromosome 22, in close proximity to two pseudogenes (*CYP2D7P* and *CYP2D8P*). Söderback and co-workers [4] developed a Pyrosequencing method to determine the *CYP2D6* copy number using the highly homologue *CYP2D8P* gene as a reference gene with two copies.

One of the first CNVs to be determined by Pyrosequencing was performed for the *KIT* gene [5] using one shared forward primer and two tailed specific reverse primers. Pyrosequencing has later been applied to several other gene copy number determinations [6–9].

The protocol described in this chapter includes all steps necessary to determine the relative copy number of highly similar genes by Pyrosequencing. As an example, the method to calculate the relative abundance of *ABCC6* and *ABCC6P1* is described [6].

2 Materials

2.1 Assay Design

1. To download DNA sequences, use a genome database, e.g., Genome Browser (<http://genome.ucsc.edu/>).
2. To align paralogous sequences, use for example Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>).
3. Design PCR and sequencing primers using PyroMark Assay design or equivalent.

2.2 Sample Preparation and PCR Amplification

1. DNA extraction kit (variable).
2. Primers for PCR amplification (one biotinylated).
3. PyroMark® PCR kit including HotStar Taq DNA polymerase (Qiagen).
4. PCR tubes or plate (0.2 mL).
5. PCR thermal cycler.

2.3 Sample Preparation for Pyrosequencing Analysis

1. PyroMark Q24 Vacuum Workstation.
2. PyroMark Vacuum Prep Filter Probe.
3. Heating block.

4. PyroMark Q24 Plate Holder.
5. Thermomixer (room temperature).
6. PCR tubes or plate (0.2 mL).
7. PCR tube caps/strips or sealing tape.
8. PyroMark Binding Buffer.
9. Streptavidin Sepharose® High Performance (GE Healthcare).
10. PyroMark Q24 Plate.
11. Primer for Pyrosequencing.
12. PyroMark Annealing Buffer.
13. PyroMark Q24 Vacuum Prep Troughs.
14. 70 % EtOH.
15. PyroMark Denaturation Buffer.
16. PyroMark Wash Buffer.

2.4 Pyrosequencing and Analysis

1. PyroMark Q24 Instrument.
2. PyroMark Gold Q24 Reagents.
3. PyroMark Q24 Cartridge.
4. PyroMark Q24 software.

3 Methods

3.1 Assay Design

1. Alignment of gene sequences:
 - (a) Download the target and paralogous DNA sequence(s) from a genome database, e.g., Genome Browser (<http://genome.ucsc.edu/>) (*see Note 1*).
 - (b) Add the downloaded DNA sequences in FASTA format to a sequence aligner, e.g., Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>).
2. Investigate the alignment results for sequence similarity and dissimilarity. Determine target and reference nucleotides (*see Note 2*).
3. Design PCR primers using the PyroMark Assay design software or equivalent. Optimal size of the amplification product should be between 80 and 200 bp. The primers must be located in sequence regions that are common for all targeted genes (Fig. 1).

One of the primers must be biotinylated and positioned in the opposite direction of the sequencing primer (*see Note 3*). The forward and reverse primers for *ABCC6* and *ABCC6PI* amplification are 5'-TGAGGGAGCCAGGCTAGA-3' and 5'-biotin-GAGGGGAAGGGAGAGATTAGC-3', respectively.

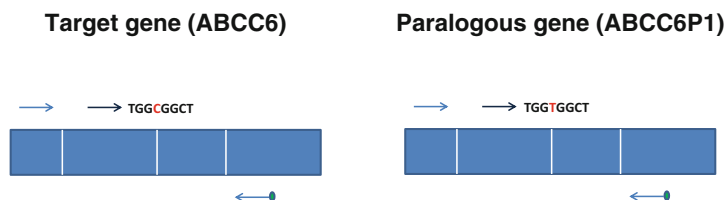


Fig. 1 Target (*ABCC6*) and paralogous (*ABCC6P1*) genes. White vertical lines indicate sequence differences between *ABCC6* and *ABCC6P1*. Note that PCR primers (blue arrow) and sequencing primers (black arrow) are located in regions with identical sequence homology

4. BLAST the PCR primers and make sure that they only amplify the targeted sequences. Avoid primer design in regions with SNPs (*see Note 4*).
5. Design the Pyrosequencing primers manually or by using the PyroMark Assay design software. The primer must be positioned in a region that is common for all targeted genes. The sequencing primer for *ABCC6* and *ABCC6P1* sequencing is 5'-GCCTGGCCCTGCCGC-3'.
6. Create an assay for Pyrosequencing using the PyroMark Q24 software. Carefully design the order of nucleotide dispensation. The target and reference nucleotides should come out as single peaks (*see Note 2*). At least one control nucleotide should be added (zero peak height is expected). A dispensation order that allows for the calculation of gene copies at different nucleotide positions is preferable and may be used as a control for the assay (Table 1 and Fig. 2).

3.2 Sample Preparation and PCR Amplification

1. Extract DNA using an appropriate protocol (*see Note 5* and, e.g., Chapter 13 for a detailed protocol).
2. Use ~1–10 ng of genomic DNA (*see Note 6*) in a 25 μ L PCR reaction together with 1 $\times$ PyroMark PCR Master Mix, 1 $\times$ CoralLoad Concentrate, and 0.2 μ M primers. The PCR cycling condition consists of a HotStar Taq DNA Polymerase activation step of 15 min at 95 $^{\circ}$ C followed by 45 cycles of 30 s at 94 $^{\circ}$ C, 30 s at the specific annealing temperature (may vary), and 30 s at 72 $^{\circ}$ C, and finally, an extension step for 10 min at 72 $^{\circ}$ C. The annealing temperature for the *ABCC6* and *ABCC6P1* PCR amplification is 60 $^{\circ}$ C.
3. Run ~5 μ L of PCR products on a 2 % agarose gel. Verify that the PCR amplicon is of correct size and that no other PCR fragments are visible on the gel. Make sure that the no-template control is negative.
4. Store PCR fragments at 4 $^{\circ}$ C for a few days or at -20 $^{\circ}$ C for a long term.

Table 1

Nucleotide dispensation order and nucleotide count for *ABCC6* and *ABCC6P1*. $C_8/$

$C_8 = ABCC6_{CN}/ABCC6_{CN} + ABCC6P1_{CN}$. If the copy number of *ABCC6*=2 and *ABCC6P1*=2 then the

expected peak height ratios between the two genes will be $2/2 + 2 = 0.5$. If the copy number of

ABCC6=2 and *ABCC6P1*=3, then the expected peak height ratio would be $2/2 + 3 = 0.40$. Note that $T_6/$

T_9 will give the ratio of *ABCC6P1* relative to the total copy numbers: $ABCC6P1_{CN}/ABCC6_{CN} + ABCC6P1_{CN}$

Nucleotide no.	1	2	3	4	5	6	7	8	9	Sequence
Nucleotide dispensation order	G	T	A	G	C	T	G	C	T	
Target gene (<i>ABCC6</i>)	0	1	0	2	1	0	2	1	1	TGGCGGCT
Paralogous gene (<i>ABCC6P1</i>)	0	1	0	2	0	1	2	1	1	TGGTGGCT

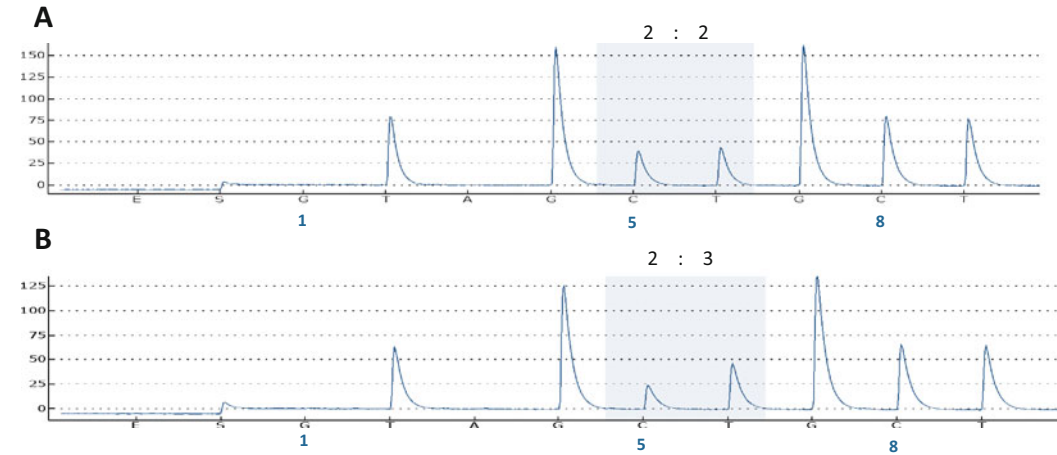


Fig. 2 Pyrogram results for the determination of the copy number ratios between *ABCC6* and *ABCC6P1*. (a) $ABCC6_{CN} = ABCC6P1_{CN} = 2$ and (b) $ABCC6_{CN} = 2$ and $ABCC6P1_{CN} = 3$. Note the reduced C_5 nucleotide peak height in (b)

3.3 Pyrosequencing

1. In the PyroMark Q24 software, create a new run with the assay designed in **step 6** of Subheading 3.1 and load the file to a USB stick. Print the instrument setup document.
2. In PCR tubes, mix 5–20 μ L of PCR product (*see Note 7*) with 40 μ L of binding buffer, 1 μ L of Sepharose beads, and water to 80 μ L total volume. Seal the tubes and incubate at room temperature with constant mixing (1,400 rpm) for 10 min.
3. During the incubation step, prepare the sequencing plate with 0.3 μ M sequencing primer in a total volume of 25 μ L annealing buffer.
4. Fill the troughs of the PyroMark Vacuum Prep workstation with 70 % ethanol, denaturation solution, washing buffer, and water, respectively.
5. Turn on the vacuum and wash the filter probes by aspirating in water for several seconds.

6. Turn the vacuum off while placing the PCR/binding mix on the workstation. Remove the sealing, turn the vacuum on again, and aspire the PCR/binding mix for about 10 s.
7. Transfer the probes to the ethanol tray for 5 s, then the denaturation solution for 5 s, and finally washing buffer for 10 s.
8. Turn off the vacuum and release the PCR product into the sequencing tray by shaking the tool from side to side.
9. Place the sequencing tray on a heating block set to 80 °C and incubate for 2 min.
10. Leave at room temperature for at least 5 min before sequencing.
11. Use the instrument setup document from **step 1** and fill the cartridge with the required volumes of enzyme, substrate, and nucleotides.
12. Run the Pyrosequencing file downloaded on the USB stick (from **step 1**).
13. Aliquots of dissolved enzyme and substrate mix can be stored at -20 °C for long-term storage.

3.4 Analysis of Results

1. When the sequencing is finished, open the run file in the PyroMark software.
2. Export the peak heights as a text file (tsv).
3. Open the text file in an Excel sheet. Calculate the ratio between the nucleotides specific for the target gene and the nucleotide specific for the paralogous gene. Compare the results to the expected peak height ratios (Table 1 and Fig. 2) (*see Note 8*).
4. Check that the height of the peaks at control nucleotides is zero.
5. PCR annealing temperature may affect the amplification of target genes, therefore control for equal amplification of all target sequences. If DNA samples with known copy number are available, check that peak height ratios are in accordance with expected results. The calculated copy number should be equal to the actual copy number (*see Note 9*).

4 Notes

1. When downloading sequencing data, be aware of the orientation of the sequences. All sequences to be aligned must be in the same reading frame.
2. Reference nucleotides are sequence variant(s) between the target and paralogous genes that may be used for the calculation of ratios between the genes. Since different nucleotides may differ slightly in peak heights, the same nucleotides should be used as reference nucleotide; for example a C peak should be compared to another C peak (Table 1). Furthermore, a

single-nucleotide difference between target and paralogous genes is optimal as di/trinucleotide repeats tend to give less precise peak height ratios.

3. To ensure that primers are designed to the region of interest, when using the PyroMark Assay Design software one may narrow down the primer search by highlighting the region of interest. Then right click on the mouse, and set forward and/or reverse primers.
4. If SNPs in the amplification primers cannot be avoided, design primers with neutral nucleotides at the SNP site, i.e., a nucleotide that will bind equally to both sequence variants.
5. Use EDTA as anticoagulant for blood samples as heparin may inhibit the PCR reaction.
6. Low DNA concentration (<1 ng) may give false results as one gene may be preferentially amplified. Especially in the case of low DNA template concentrations, it is recommended to run parallel reactions. When DNA template concentration is too low, an unspecific amplification may occur and duplicate experiments will usually vary a great deal. Such samples should be excluded from further analysis.
7. Usually 5–10 µL of PCR product is enough for Pyrosequencing, but for low-intensity bands on agarose gel, use 20 µL.
8. Be aware that the calculations of gene copy numbers of target genes are based on the relative abundance of the paralogous gene(s). For example an equal ratio between target and paralogous genes may indicate gene copy numbers from 1, 2, 3, etc.
9. If control samples with known copy number are not available, plasmid constructs of the target and paralogous genes can be made and tested for copy number changes using different ratios of these constructs as template for PCR.

References

1. Stankiewicz P, Lupski JR (2010) Structural variation in the human genome and its role in disease. *Annu Rev Med* 61:437–455
2. Schaeffeler E, Schwab M, Eichelbaum M et al (2003) CYP2D6 genotyping strategy based on gene copy number determination by TaqMan real-time PCR. *Hum Mutat* 22:476–485
3. Owen RP, Sangkuhl K, Klein TE et al (2009) Cytochrome P450 2D6. *Pharmacogenet Genomics* 19:559–562
4. Soderback E, Zackrisson AL, Lindblom B et al (2005) Determination of CYP2D6 gene copy number by pyrosequencing. *Clin Chem* 51: 522–531
5. Pielberg G, Day AE, Plastow GS et al (2003) A sensitive method for detecting variation in copy numbers of duplicated genes. *Genome Res* 13:2171–2177
6. Kringen MK, Stormo C, Grimholt RM et al (2012) Copy number variations of the ATP-binding cassette transporter ABCB6 gene and its pseudogenes. *BMC Res Notes* 5:425
7. Liu Z, Obenauf AC, Speicher MR et al (2009) Rapid identification of homologous recombinants and determination of gene copy number with reference/query pyrosequencing (RQPS). *Genome Res* 19:2081–2089
8. Liu Z, Schneider DL, Kornfeld K et al (2010) Simple copy number determination with reference query pyrosequencing (RQPS). *Cold Spring Harb Protoc* 9:5491
9. Setty P, Gessi M, Waha A et al (2011) Sensitive determination of BRAF copy number in clinical samples by pyrosequencing. *Diagn Mol Pathol* 20:148–157

Free ebooks ==> www.Ebook777.com

Chapter 10

Prenatal Diagnosis of Chromosomal Aneuploidies by Quantitative Pyrosequencing®

Hui Ye, Haiping Wu, Yunlong Liu, Bingjie Zou, Tomoharu Kajiyama, Hideki Kambara, and Guohua Zhou

Abstract

One of the major reasons for pregnant women to ask for prenatal diagnosis is to detect fetal chromosomal aneuploidies. Analysis of allele ratios of SNPs has been used for prenatal detection of fetal aneuploidies using MALDI-TOF mass spectrometry (MS). However, quantitative SNP genotyping by MALDI-TOF MS is challenging. To obtain a better quantification of allelic ratios, a Pyrosequencing® protocol for SNP genotyping has been developed to perform prenatal diagnosis of aneuploidies.

To avoid the laborious process and risk of cross-contamination brought in by DNA extraction procedures, a PCR assay, which can amplify DNA directly from cells in amniotic fluid, has been developed. Pre-amplification steps such as cell enrichment and heating are required to obtain sufficient amounts of amplification products.

In this chapter, SNPs on chromosome 21 are used to detect trisomy 21 as an example of aneuploidy by quantifying the allele ratio using Pyrosequencing. Primer selection for PCRs and Pyrosequencing reactions, optimization of nucleotide dispensation orders, establishment of cutoff values for trisomy 21, and interpretation of data are all factors essential for a successful diagnosis and are discussed in detail herein.

Key words Pyrosequencing®, Amniotic fluid, Heterozygote, Trisomy21, SNP, Prenatal diagnosis, Aneuploidy

1 Introduction

Detection of fetal chromosomal aneuploidies is one of the major tasks in prenatal diagnosis. Karyotyping, which has for many years been the gold standard of prenatal diagnosis, requires lengthy procedures of culturing amniotic fluid and chorionic villus cells before a result can be issued, exacerbating parental stress during this process. However, with PCR-based methods, a rapid diagnosis can be achieved even by using only a small amount of specimen for the analysis [1–3].

Heterozygous SNPs on common trisomic chromosomes could be ideal diagnostic biomarkers. Given chromosome 21 as an example, a euploid fetus inherits one chromosome from each parent, which means

his/her paired chromosome 21 should have many polymorphisms with an allelic ratio of 1:1. However, a fetus with trisomy 21 has one more copy of chromosome 21, and thus, the allelic ratio of a heterozygous SNP on chromosome 21 will be 2:1 or 1:2.

Quantitative detection of allelic ratios for prenatal diagnosis was first proposed by Lo et al. [4, 5]. The detection was performed by means of a single base-extension assay coupled with MALDI-TOF mass spectrometry (MS), and the quantification was achieved by calculating the relative yield of each allele peak in a MS spectrum of an SNP. However, the quantitative performance of MALDI-TOF MS was not satisfactory [5, 6].

Pyrosequencing is a sequencing-by-synthesis method, which is based on the bioluminometric detection of inorganic pyrophosphates (PPi) coupled with multiple enzymatic reactions (*see* Chapter 1). In a Pyrogram® the relative intensity of each peak is proportional to the number of incorporated nucleotides. As only one nucleotide is incorporated at a time, the intensity of each peak is proportional to the amount of template in the reaction. Therefore, two peaks with equal height should appear in a Pyrogram for an SNP if it is not part of a homopolymer, which might distort the Pyrogram. Consequently, two peaks with the ratio of 1:2 or 2:1 would be observed for a trisomy 21 fetus when the SNP of interest is located on chromosome 21.

A conventional Pyrosequencing protocol generally includes DNA extraction from raw samples, PCR amplification of target sequences, single stranded DNA preparation, and eventually the Pyrosequencing reactions. In such protocols, DNA extraction, which serves as a purification process, is needed prior to amplification due to the negative effects of components in the raw samples. However, DNA extraction is costly and laborious and most importantly, increases the risk of cross-contamination and loss of nucleic acids. Hence, it is preferred to employ an amplification strategy which directly uses raw samples (e.g., blood or amniotic fluid).

One of the components that inhibit amplification in raw samples is positively charged proteins. These proteins interact with DNA and hinder the binding of the polymerase to DNA. Hence, the pH of the PCR buffer is increased by adding $(\text{NH}_4)_2\text{SO}_4$ to change the net charge of proteins, thus exposing DNA to the polymerase [7]. To further improve the yield of the amplification products, this protocol increases the amount of DNA template by centrifugation and preheating of cells in amniotic fluid.

Using this simplified extraction and amplification protocol, several SNP loci can be analyzed by Pyrosequencing. Subsequently, genotypes can be determined by calculating the peak height ratio of the two alleles in a Pyrogram. Since the allelic ratio of each SNP may fluctuate with its sequence, cutoff values of the allelic ratio for each SNP have to be determined. Furthermore, clinical samples will be sequenced and diagnosed according to the experimentally determined cutoff values.

2 Materials

2.1 DNA Extraction

1. Proteinase K.
2. Chelex 100.
3. NaCl (0.9 %).

2.2 PCR Amplification

1. PCR primers (designed using the PyroMark Assay 2.0).
 2. Biotinylated primers.
 3. *Taq* DNA polymerase.
 4. dNTPs (10 mmol/L).
 5. HpH Buffer: 100 mmol/L Tris-HCl, 50 mmol/L KCl, 10 mmol/L (NH₄)₂SO₄, pH 9.3–9.5.
 6. Thermocycler: Eppendorf Mastercycler® nexus.
- All reagents should be stored at –20 °C.

2.3 Single-Stranded DNA (ssDNA) Preparation

1. Streptavidin Sepharose® High Performance Beads.
2. 1× binding and wash buffer: 5 mmol/L Tris-HCl, 0.5 mmol/L EDTA, 1.0 mol/L NaCl, pH 7.5.
3. Denaturing solution: 0.1 mol/L NaOH.
4. Annealing buffer: 4 mmol/L Tris-HCl, 2 mmol/L MgCl₂, 5 mmol/L NaCl, pH 7.5.
5. Sequencing primer.

2.4 Pyrosequencing Reaction

- 40 µL of pyrosequencing mixture contains:
1. 0.1 mol/L Tris-HAc, pH 7.7.
 2. 2 mmol/L EDTA.
 3. 10 mmol/L Mg(Ac)₂.
 4. 0.1 % BSA.
 5. 1 mmol/L DTT.
 6. 2 µmol/L adenosine 5'-phosphosulfate (APS, Sigma-Aldrich; cat. no. A5508).
 7. 0.4 g/L PVP.
 8. 0.4 mmol/L D-luciferin (Sigma-Aldrich; cat. no. L9504-5MG).
 9. 2 µmol/L ATP sulfurylase (Sigma-Aldrich; cat. no. A8957).
 10. 1.6 U/mL Apyrase-VII (Sigma-Aldrich; cat. no. A6535).
 11. 18 U/mL Exo-Klenow fragment (Promega; cat. no. M2181).
 12. An appropriate amount of QuantiLum® Recombinant Luciferase (Promega; cat. no. E1702), usual concentration: 1.5 µg/40 µL.
 13. Portable bioluminescence analyzer (Hitachi Ltd., Japan).

3 Methods

3.1 SNP Selection

Taking trisomy 21 as an example, SNPs on chromosome 21 are targeted for diagnosis. The copy number of chromosome 21 will be determined by calculating the peak height ratio of the two alleles in a Pyrogram. Therefore, the selection of SNP is critical for the following steps.

An ideal SNP should meet the following criteria:

1. It is located on chromosome 21.
2. It has a relatively high heterozygosity (above 0.4 is recommended) within the population investigated.
3. No linkage exists between any two SNPs when a panel of SNPs is selected.
4. The population coverage of all selected SNPs should be above 90 %.
5. The sequences in direct vicinity of a targeted SNP should have less than two identical nucleotides.

3.2 Design of PCR Primers and Sequencing Primers

Pyrosequencing might fail due to a low signal-to-noise ratio or a high nonspecific peak. It is crucial to design PCR primers and sequencing primers in order to achieve accurate and reliable Pyrosequencing results. The PyroMark Assay Design 2.0 software from Qiagen is recommended to design primers. Below, some guidelines are listed to improve primers generated by the software:

1. Avoid self-complementarity of sequencing primers, because this is the main cause of low signal intensity and nonspecific signals in a Pyrogram.
2. The first nucleotide incorporated in a Pyrosequencing reaction does not necessarily need to be within the SNP site.
3. The length of the amplification product should be between 70 and 150 base pairs. Longer amplification products may result in complex secondary structures and produce nonspecific signals in Pyrograms, while short amplification products may be difficult to distinguish from primer dimers in the 2 % agarose gel electrophoresis.
4. The sense and antisense sequence near an SNP should be investigated to choose the optimum strand for sequencing avoiding homopolymeric regions near the SNP and resulting in an “easy-to-translate” Pyrogram. After strand selection, the respective amplification primer should be modified with a biotin group.
5. Verify the specificity of primers, the absence of the amplification of pseudo genes and segmental duplications by blasting PCR primer sequences using NCBI Primer Blast (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>).

6. The detection of a “C/T”-type SNP is preferred compared to an “A/G”-type SNP, because “dATPαS” generally produces a slightly higher signal compared to the other three nucleotides. Considering the importance of quantifying the allelic ratio in this protocol, a “C/T”-type SNP is less biased and should therefore be preferred.

3.3 Sample Processing

Generally, an amniotic fluid sample in the clinics needs to be analyzed as fast as possible in order to reduce the waiting period before the result is issued. Therefore, PCR amplification directly from the amniotic fluid sample is preferable. Amniotic fluid sample can be kept at 4 °C for as long as 3–6 months. However, it is not recommended to maintain raw samples for such a long run, especially considering the possibility of bacterial or fungal contamination of the specimen. Therefore, a protocol for DNA extraction and storage from amniotic fluid samples is provided here.

3.3.1 For Prenatal Diagnosis

1. Collect at least 1 mL of amniotic fluid, and store at 4 °C before use.
2. Centrifuge 1 mL of amniotic fluid in a 1.5 mL microcentrifuge tube at $5,000 \times g$ for 10 min, then remove 900 μ L of the supernatant, mix the remainder evenly (*see Note 1*).
3. Transfer a certain amount of the concentrated amniotic fluid (depending on the number of targeted SNPs) to a 200 μ L tube and heat at 94 °C for 15 min. Generally, 1–3 μ L of the treated amniotic fluid could be used in a 50 μ L PCR reaction.
4. Keep the treated amniotic fluid on ice before use.

3.3.2 For Sample Storage

1. Centrifuge 1 mL of amniotic fluid in a 1.5 mL microcentrifuge tube at $5,000 \times g$ for 10 min. Remove all the supernatant.
2. Suspend the pellet in 20 μ L of 0.9 % NaCl, and mix evenly.
3. Transfer the suspension to 250 μ L 5 % Chelex 100 and add 10 μ L of 20 mg/mL Proteinase K.
4. Incubate at 56 °C for 30 min, then vortex at a high speed for 5–10 s.
5. Incubate at 98 °C for 8 min, then vortex at a high speed for 10 s.
6. Centrifuge the mix at $10,000 \times g$ for 3 min. Transfer the supernatant (approximate 200 μ L) to a new tube, avoiding the carryover of any sediments from the bottom. The extracted DNA can be stored at –20 °C.
7. Generally, 10 μ L of the extracted amniotic fluid DNA can be used in a 50 μ L PCR reaction.

3.4 PCR Amplification

Since two kinds of templates (raw DNA from amniotic fluid and extracted DNA) might be used after sample processing, two PCR protocols are provided.

3.4.1 For Raw DNA from Amniotic Fluid

1. Set up a 50 μL reaction containing 5.35 μL of “HpH Buffer,” 200 $\mu\text{mol/L}$ each of dATP, dCTP, dGTP, and dTTP, 0.4 $\mu\text{mol/L}$ of each primer, and 2.5 U of *Taq* Polymerase (*see Note 2*).
2. Initiate the reaction by incubation at 94 °C for 5 min, followed by 35 cycles of 94 °C for 30 s, 55–60 °C for 30 s, 72 °C for 30 s, and a final extension at 72 °C for 7 min.

3.4.2 For Extracted DNA

1. Set up a 50 μL reaction containing 5 μL of 10 $\times$ buffer (purchased with the polymerase), 2 mmol/L of MgCl_2 , 200 $\mu\text{mol/L}$ each of dATP, dCTP, dGTP, and dTTP, 0.4 $\mu\text{mol/L}$ of each primer, 1.25 U of *Taq* Polymerase.
2. Initiate the reaction by incubation at 94 °C for 5 min, followed by 35 cycles of 94 °C for 30 s, 55 °C for 30 s, 72 °C for 30 s, and a final extension at 72 °C for 7 min.

3.5 Single-Stranded DNA (ssDNA) Preparation

1. Suspend 5 μL of Sepharose beads in 50 μL 2 $\times$ binding and wash buffer.
2. Mix 50 μL of PCR product with the suspended beads (*see Note 3*). Incubate while rotating/shaking at 37 °C for 15 min.
3. Centrifuge and remove the supernatant carefully, avoiding the loss of beads.
4. Wash the beads with 100 μL of H_2O twice.
5. Denature double-stranded amplification products in 25 μL of 0.1 mol/L NaOH for 5 min. Keep the beads suspended in the solution.
6. Centrifuge and remove the supernatant carefully (*see Note 4*).
7. Wash the beads immediately twice with 50 μL of 1 $\times$ Annealing buffer.
8. Suspend the beads in 10 μL of 1 $\times$ Annealing buffer. Keep it at 4 °C before use (*see Note 5*).
9. Add 1 μL of sequencing primer (10 pmol/ μL) to the ssDNA, and mix evenly.
10. Anneal the sequencing primer to the immobilized biotinylated strand used as sequencing template by incubation at 80 °C for 5 min and 25 °C for 10 min.

3.6 Pyrosequencing

We use a portable bioluminescence analyzer for Pyrosequencing. This device has a portable size of 140 mm (W) $\times$ 158 mm (H) $\times$ 250 mm (D), and uses a photodiode array to detect photon signals, and four separate capillaries to dispense small amounts of dNTPs into the reaction chamber.

The volume of a Pyrosequencing reaction was 40 μL , containing 0.1 mol/L Tris–acetate (pH 7.7), 2 mmol/L EDTA, 10 mmol/L magnesium acetate, 0.1 % BSA, 1 mmol/L dithiothreitol (DTT),

2 $\mu\text{mol/L}$ adenosine 5'-phosphosulfate (APS), 0.4 mg/mL PVP, 0.4 mmol/L D-luciferin, 2 $\mu\text{mol/L}$ ATP sulfurylase, an appropriate amount of QuantiLum® Recombinant Luciferase (1.5 $\mu\text{g}/40\ \mu\text{L}$), 18 U/mL *Exo*-Klenow fragment, and 1.6 U/mL Apyrase-VII. Each of the dNTPs was added in the reservoir of the micro-dispenser separately, and dispensed into Pyrosequencing reaction in an assigned order.

3.7 Establishment of Cutoff Values for Allelic Ratios

The expected allelic ratio of an SNP in a Pyrogram is 1 for a euploid fetus, and 2 or 0.5 for a trisomic fetus. However, this is not always the case. Homopolymeric sequences around the SNP locus may alter the read-out of the allelic ratio in a Pyrogram. For instance, the signal intensity ratio of “C/T” for SNP rs767055 is 1 for a euploid heterozygous (Fig. 1a), which is equal to its allelic ratio. However, the signal intensity ratio of “C/T” at rs914232 for a euploid heterozygous is 3 (Fig. 1b) (*see Note 6*).

Most importantly, the signal intensity ratio of two nucleotides may vary for a targeted sequence. Although this variation is not a problem for SNP typing, it may lead to a bias in the determination of the precise allelic ratio. It is therefore of importance to determine a cutoff value for each SNP defined using euploid heterozygotes. Hence, samples, which have been tested by karyotyping as euploid should be employed in this protocol to establish cutoff values (*see Note 7*).

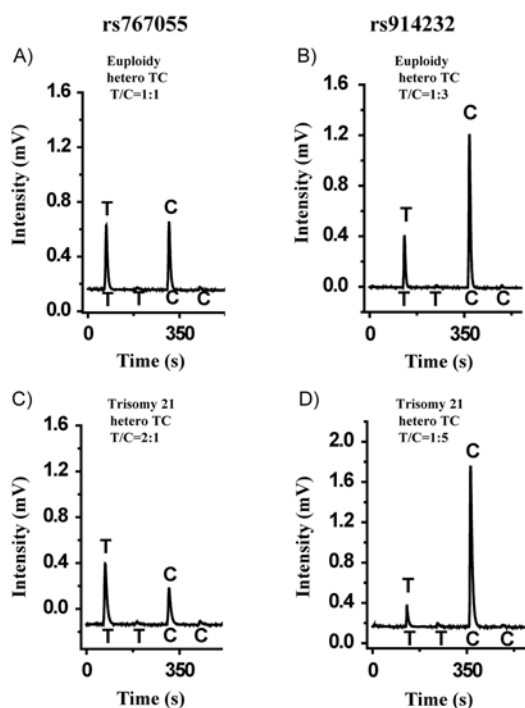


Fig. 1 Pyrograms of heterozygotes for rs767055 ((a) euploidy; (c) trisomy) and rs914232 ((b) euploidy; (d) trisomy)

1. Before Pyrosequencing each SNP, it is highly recommended to optimize the dNTP dispensation orders to obtain accurate quantitative Pyrosequencing results. For example, the sequence near rs914232 is CTC/TCTTCC. If dNTPs are added in the order of “TCAG,” the signal intensity ratio of “C/T” will be 3:1 for euploid heterozygotes, and 5:1 for trisomic heterozygotes. The differences between a euploid and a trisomic sample could be easily determined from the Pyrograms. However, if dNTPs would be added in an order of “CTAG,” the signal intensity ratio of “C/T” will be 1:1.5 for euploid heterozygotes, and 1:1.25 for trisomic heterozygotes. Under these circumstances, Pyrograms are not obviously different between the two genotypes.
2. Analyze SNPs in at least 30 euploid samples by Pyrosequencing®, and record the corresponding signal intensity of alleles for each heterozygous SNP (*see Note 8*).
3. Calculate the height ratio of the two allele-specific peaks in a Pyrogram.
4. Calculate the cutoff value of each SNP as the mean heterozygote allelic ratio $\pm [1.96 \times \text{SD}]$.
5. With the intervals of cutoff value determined, verify the percentage of detected allelic ratios that fall within these intervals. If the percentage is close to 95 %, the cutoff values can be used to identify trisomy 21 fetuses.
6. During this process, a panel of SNPs could be selected based on the prevalence of their heterozygosity. The population coverage should be close to 100 %.

3.8 Genotyping and Diagnosing Clinical Samples

With the cutoff values of each SNP defined, clinical samples can be diagnosed. The criteria to determine if a sample is trisomic or not are as follows:

1. For each sample, at least two heterozygous SNPs would be sufficient for an accurate diagnosis.
2. The detected allelic ratios of all heterozygotes for a trisomy should be out of the confidence intervals, indicating that these samples are significantly different from euploid samples (as shown in Fig. 1c, d).
3. Z-scores could be employed to evaluate the confidence level of detected results. Calculate Z-scores according to the formula: “Z-score = (the detected allelic ratio of a sample – the mean allelic ratio of euploid heterozygotes) / SD of the mean allelic ratio of heterozygotes.”
4. With Z-score higher than 2.576 or lower than –2.576, the degree of confidence level to call a sample trisomy is above 99 %; with Z-score higher than 1.96 or lower than –1.96, the degree of confidence level to call a sample trisomy is above 95 % (*see Note 9*).

4 Notes

1. The reason for keeping 100 μ L of amniotic fluid is to maintain suitable environment for cells in case one may keep them at 4 °C for a period of time, and leave a certain amount of volume for multiple SNP detections.
2. Add the treated amniotic fluid to the reaction without any disturbance.
3. Make sure the volume ratio of PCR product to suspended beads is 1.
4. The supernatant could be transferred to another tube, neutralized with 1 mol/L HCl, and then serve as sequencing template as well.
5. Immobilized single-stranded DNA could be kept at 4 °C for 1–3 months, as long as the tube is sealed by Parafilm.
6. Sequences of primers used for genotyping rs767055 and rs914232 are listed in Table 1. As shown in the table, rs767055 has no homopolymeric sequence near SNP, while rs914232 has.
7. At the same time of establishing cutoff values, the frequency of heterozygosity of an SNP in the population could also be validated, as information from a database may not fully represent the heterozygosity in a population. Although methods such as allele-specific PCR have been widely applied for SNP genotyping and the process is less laborious compared to sequencing, it is less accurate and may generate false-positive results. Therefore, results from sequencing are much more reliable.
8. As a small amount of PPi (due to the decomposition of dNTPs) may exist in dNTPs, each dNTP is dispensed twice,

Table 1
Sequences of primers used for genotyping rs767055 and rs914232

SNP	Sequence (5' → 3')	Amplicon size (bp)	Sequence to be analyzed
rs767055	PCR-F ^a : GTG CGA AAA GGA AGT GAT GCC PCR-R ^a : Biotin-AAT ATC CCA TTG GTG CCT AGT GTT Pyro-P ^b : AAT ATC CAT GAA AAA GCC AA	146	C/TAGCGTG
rs914232	PCR-F: Biotin-CCC AGC CAG CTT CTC TAT TTT TA PCR-R: TGT CTT CTG TTT TTG TCT CTC TCC A Pyro-P: TTC ACA CCA GCA CAA GCT	124	C/TCTTCCA

^aPCR-F and -R: paired primers used for PCR

^bPyro-P: primer used for Pyrosequencing

and a peak height is then calculated by subtracting the noise at the second dNTP dispensation from the signal at the first dNTP dispensation.

9. Low signal intensity and high background signals may result in biased *Z*-scores. If two heterozygous SNP of a sample show ambiguous results, check the quality of the Pyrogram and repeat the assay. When repeating the assay, more DNA template could be added to PCR, and more immobilized ssDNA template could be added to a Pyrosequencing[®] reaction.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (project number: 21275161 and 21005088) and the National Key Basic Research Program of China (2014CB744501).

References

1. Song Q, Wei G, Zhou G (2014) Analysis of genetically modified organisms by pyrosequencing on a portable photodiode-based bioluminescence sequencer. *Food Chem* 154:78–83
2. Song Q, Yang H, Zou B et al (2013) Improvement of LATE-PCR to allow single-cell analysis by pyrosequencing. *Analyst* 138:4991–4997
3. Chen Z, Fu X, Zhang X et al (2012) Pyrosequencing-based barcodes for a dye-free multiplex bioassay. *Chem Commun* 48:2445–2447
4. Tsui NB, Chiu RW, Ding C et al (2005) Detection of trisomy 21 by quantitative mass spectrometric analysis of single-nucleotide polymorphisms. *Clin Chem* 51:2358–2362
5. Lo YM, Tsui NB, Chiu RW et al (2007) Plasma placental RNA allelic ratio permits noninvasive prenatal chromosomal aneuploidy detection. *Nat Med* 13:218–223
6. Trewick AL, Moustafa JS, de Smith AJ et al (2011) Accurate single-nucleotide polymorphism allele assignment in trisomic or duplicated regions by using a single base-extension assay with MALDI-TOF mass spectrometry. *Clin Chem* 57:1188–1195
7. Bu Y, Huang H, Zhou G (2008) Direct polymerase chain reaction (PCR) from human whole blood and filter-paper-dried blood by using a PCR buffer with a higher pH. *Anal Biochem* 375:370–372

Chapter 11

HLA-B and HLA-C Supratyping by Pyrosequencing®

Irene Vanni, Elisabetta Ugolotti, Patrizia Larghero, and Roberto Biassoni

Abstract

Usually, HLA typing has been performed either by serology-based typing incubating a panel of known anti-HLA antibodies with viable lymphocytes of unknown HLA type or by molecular typing including medium-resolution HLA typing by Sequence Specific Oligonucleotide Probes (SSOP) or high-resolution HLA typing by Sequence Based Typing (SBT). Traditionally, HLA antigens have been defined using serological techniques, but these methods have several disadvantages, such as low resolution, the requirement for viable cells, and cell surface expression of HLA molecules. HLA type screening methods are categorized as low, medium, and high resolution, and only sequencing-based typing methods provide the highest resolution and are considered the gold standard for HLA typing.

Among the HLA SBT based-methods, the Pyrosequencing® technique is an extremely versatile and accurate real-time sequencing technique with some advantages compared to classic Sanger method.

Here, we describe a quick and inexpensive method that allows through the use of Pyrosequencing subtyping of HLA class I molecules, into HLA-Bw6, -Bw4 I80, or -Bw4 T80 and HLA-C1, or -C2 groups. In particular, this analysis is focused on the amino acids around residue 80. This method demonstrated good sensitivity, specificity, and reproducibility. Using a quantitative allele acquisition mode, the method provides accurate sequence information required for the definition of heterozygous and/or homozygous samples.

Key words HLA, MHC, Pyrosequencing®, Supratyping, Real-time sequencing

1 Introduction

Alloreactivity not only is fundamental for the activation of T cell receptors (TCR) and thus for the regulation of adaptive immune responses, but plays also a major role in the regulation of the innate immune system [1–3]. Natural Killer (NK) cells are part of an intricate network of cellular interaction and cytokine release that shape the adaptive cellular response.

The function of NK cells and a fraction of CD8+ T lymphocytes are regulated by an array of HLA class I-specific killer immunoglobulin-like receptors (KIR) able to deliver either inhibitory or activating signals that like TCRs recognize only a partially overlapping area on human leukocyte antigen (HLA) class I molecules [4].

Inhibitory KIRs are able to recognize selected groups of HLA class I molecules and to sense cells that do not express or express inadequate amounts of MHC class I molecules, resulting in their lysis mechanism known as “missing self” [5].

The $\alpha 1$ -helix region of the HLA class I heavy chain is defining the NK cell specificity and thus the alloreactivity, and in particular the amino acid 80 is known to be central for the pattern recognition and avidity of KIR/HLA class I interactions. Thus, HLA-B molecules can be divided into two mutually exclusive HLA-Bw4 and HLA-Bw6 variants defined by amino acid sequence variation at residues 77–83 of the HLA α -1 domain, while HLA-C may be subdivided in C1 or C2 group of HLA-C alleles investigating the area around residue 80 [6–11].

The *KIR3DL1* locus encodes a specific receptor for HLA-B alleles sharing the Bw4 public epitope and among them, the Bw4-alleles characterized by Ile-80 residue give an optimal inhibitory signal [12–14], while the presence of Thr-80 has been responsible for weaker inhibition [14]. No KIR specificity has been associated with the Bw6 epitope recognition [1].

KIR2DL1 is specific for HLA-C alleles of group 2, all characterized by Lys-80, while KIR2DL2 and KIR2DL3 recognize essentially group 1 HLA-C alleles containing Asn-80 [6–8]. Thus the knowledge of the supra-typing for HLA class I might be indicative of a group of alleles that could be involved in the NK alloreactivity [15, 16].

2 Materials

All steps require:

1. Pipettes P10, P20, P200, P1000, and multichannel pipettes.
2. Filter tips only of RNase, DNase, DNA and PCR inhibitors free.
3. RNase, DNase, DNA and PCR inhibitors free tubes.
4. Disposable gowns and gloves (*see Note 1*).
5. Biocontainment hoods to prevent DNA contamination of laboratory and samples.

2.1 DNA Extraction

1. PureLink Genomic DNA Mini Kit (Life Technologies) or Magratation System 12GC (PSS Bio Instruments).
2. Heating block.
3. Microcentrifuge and vortex.

2.2 Estimation of DNA Concentration

1. Spectrophotometer for small volumes (e.g., NanoDrop 1000, Thermo Fisher Scientific or Biospectrometer, Eppendorf).
2. Pipette P10.

**2.3 Locus-Specific
Polymerase Chain
Reaction (PCR)
and Nested
Polymerase Chain
Reaction (Nested PCR)**

1. 10 mM dNTP Mix, PCR Grade.
2. Platinum Taq DNA Polymerase KIT (Life Technologies).
3. DNA oligonucleotide primers. Desalted primers should be dissolved in nuclease-free water at 100 μ M (stock concentration) and conserved at -20°C . The primers sequences and their working concentration are shown in Table 1.
4. Nuclease-free water.
5. Biotinylated DNA oligonucleotide primers containing a 5' biotin-modification, HPLC-purified, dissolved in nuclease-free water at 100 μ M (stock concentration).

**2.4 Agarose Gel
Electrophoresis**

1. Agarose D-1 Low EEO.
2. Ethidium bromide solution (10 mg/mL).
3. TrackIt X174rf DNA/Hae III Fragments (Life Technologies) or equivalent.

Table 1
Locus specific and nested PCR primers

Primers	Sequence	Target	Final concentration (nM)
CEX2 Frw [18]	CTC TGS GGA GAG GAG GAR CGA GG	Locus HLA-C	1,000
CEX2 Rev [18]	TGA AAC CGG GTA AAG GYG ACT		1,000
5Cln1-61 [19]	AGC GAG GKG CCC GCC CGG CGA		400
3Cln3-12 [19]	GGA GAT GGG GAA GGC TCC CCA CT		400
biot-HLA-C Frw	Biot-GMG CCG TGG GTG GAG CA	Nested HLA-C	40
HLA-C Rev	CCG GCC TCG CTC TGG TTG TAG T		40
HLA-C SEQ	TGT AGT AGC CGC GCA G	HLA-C sequencing	800
BEX2 Frw [18]	GGG AGG GAA ATG GCC TCT	Locus HLA-B	1,500
BEX2 Revb [18]	GAG GGT YTG GGC GGG TC		400
5Bln1-57 [19]	GGG AGG AGC GAG GGG ACC RCA G		400
3Bln3-37 [19]	GGA GGC CAT CCC CGG CGA CCT AT		400
biot-HLA-B Frw	Biotin-GAG GGG CCG GAR TAT TGG	Nested HLA-B	100
HLA-B Rev	GGC CTC GCT CTG GTT GTA G		200
HLA-B SEQ	CTC GCT CTG GTT GTA GT	HLA-B sequencing	800

4. Loading buffer dye (6×): 0.25 % bromophenol blue, 0.25 % xylene cyanol FF, 30 % (v/v) glycerol.
5. 10× TAE Buffer: 400 mM Tris–acetate, 10 mM EDTA.
6. ChemiDoc MP system (Bio-Rad) or equivalent.
7. Pipettes P10 and P20.

2.5 Pyrosequencing

1. Streptavidin Sepharose HP, 5 mL (GE Healthcare Bio-Sciences).
2. PyroMark® Binding Buffer (Qiagen).
3. PyroMark Annealing Buffer (Qiagen).
4. PyroMark Denaturation Solution (Qiagen).
5. PyroMark Washing Buffer 10× (Qiagen).
6. PyroMark Gold Q96 Reagents (5 × 96) (Qiagen).
7. Nuclease-free water.
8. Absolute ethanol.
9. ClearSeal film for qPCR.
10. Microplate shaker instrument.
11. PSQ 96 Reagent Cartridge (Qiagen).
12. Vacuum Prep Tool Troughs (Qiagen).
13. Vacuum Prep Tool Filter Probes (Qiagen).
14. PSQ 96Plate Low (Qiagen).
15. 96-well PCR plate (full skirt)
16. PyroMark Vacuum Prep workstation (Qiagen).
17. PyroMark ID instrument (Qiagen).
18. Microcentrifuge.
19. Vortex.

3 Methods

3.1 DNA Sample Preparation

Genomic DNA suitable for the Pyrosequencing approach can be extracted starting from a pellet of $1\text{--}2 \times 10^7$ peripheral blood mononuclear cells (PBMC) (*see Note 1*) or directly from freshly drawn blood (200 μL) using different protocols that yield a high molecular weight genomic DNA. For DNA extraction from PBMCs we use the PureLink® Genomic DNA Mini Kit based on a DNA binding silica membrane, and from blood, especially for high-throughput protocols and to avoid cross-contamination, we use the automated Magstration System 12G, which uses a purification protocol based on magnetic silica particles. DNA Extraction is carried out according to the manufacturer's protocols [17].

Table 2
HLA-C locus specific PCR mix

Reagent	μL	[Concentration] _f
10× Buffer	5	
10 mM dNTP mix	1	0.2 mM
50 mM MgCl ₂	1.5	1.5 mM
Platinum Taq (5 U/ μL)	0.5	2.5 U
CEX2 FRW (10 μM)	5	1,000 nM
CEX2 REV (10 μM)	5	1,000 nM
5Cln1-61 (10 μM) [19]	2	400 nM
3Cln3-12 (10 μM) [19]	2	400 nM
Total volume mix	22	

DNA concentration is determined reading absorbance at 230, 260, 280, and 320 nm on a classical UV-Spectrophotometer or using small volume sample reader.

3.2 HLA-C Template Preparation

3.2.1 HLA-C Locus Specific PCR

- All reagents have to be dispensed in an eight-well microtube (0.2 mL) strip.
- Use for each sample 22 μL of a PCR mix containing the reagent as described in Table 2 (*see* **Note 2**).
- Add a volume containing 50 ng of DNA and adjust to 50 μL with nuclease-free distilled/deionized water.
- Cap the strip and start PCR cycling as follow: 3 min denaturation at 95 °C, 5 cycles of 15 s denaturation at 95 °C, 15 s at 62 °C for primer annealing and 1 min elongation at 72 °C, followed by 26 cycles of 15 s denaturation at 95 °C, 15 s at 58 °C for primer annealing and 1 min elongation at 72 °C and 4 cycles of 15 s denaturation at 95 °C, 15 s at 56 °C for primer annealing and 1 min elongation at 72 °C. The program is terminated with a final extension step for 7 min at 72 °C.

3.2.2 HLA-C Nested PCR

- Dispense for each sample 24 μL of the reagent mix (Table 3) containing the biot-HLA-C Frw and HLA-C Rev primers (both at 40 nM, Table 1) (*see* **Note 2**).
- Add 1 μL of the HLA-C locus specific PCR product in an eight-well microtube (0.2 mL) strip
- Cap the strip and start PCR cycling as follow: 2 min denaturation at 95 °C, 40 cycles of 15 s denaturation at 95 °C, 15 s at 60 °C for primer annealing and elongation, followed by final extension step for 5 min at 72 °C.

Table 3
HLA-C nested PCR mix

Reagent	μL	[Concentration] _f
10× Buffer	2.50	
10 mM dNTP mix	0.50	0.2 mM
50 mM MgCl ₂	0.75	1.5 mM
Platinum Taq (5 U/ μL)	0.25	1.25 U
Biot-HLA-C Frw (10 μM)	0.1	40 nM
HLA-C Rev (10 μM)	0.1	40 nM
H ₂ O	19.8	
Total volume mix	24	

	1 Parking position water
3 Denaturation solution	5 Water
2 Ethanol	4 Washing buffer
PCR plate	PSQ plate

Fig. 1 PyroMark Vacuum Prep Workstation Troughs position

3.3 Setting

Up the Pyrosequencer

3.3.1 Use the SNP
Run Mode of the
Pyrosequencing PyroMark
ID (Qiagen) Instrument

- In the SNP run set-up window, add in the Instrument Parameters the numerical code of the PSQ 96 Reagent Cartridge.
- Select and activate the sample containing wells (the active wells turn yellow).
- In the “Set up” submenu window choose the Simplex Option.
- “ID Entry” defines the name of the analysis-file (i.e., HLA-C supratyping).
- Select with the graphic pen all the wells containing samples.
- Set the nucleotide dispensation order as: CgT ACT CgC AgC T.
- Set the sequence to be analyzed: [g/T] CTC gCA g[C/T]g.

3.3.2 Setting
the PyroMark Vacuum Prep
Workstation (Fig. 1)

- Prepare 180 mL of 70 % ethanol (v/v) and pour in Trough in position 2.
- Prepare 180 mL of 1× PyroMark Washing Buffer solution in Trough 4.

- (c) Pour in trough 3 120 mL of ready to use Denaturation Solution.
- (d) Trough 1 and 5 contain each 180 mL of deionized water.
- (e) Set the microplate mixer to 1,200 rpm at room temperature.

3.3.3 Single Stranded Template Preparation

- (a) Prepare for each sample 55 μ L of Binding Mix (*see Note 3*): add 36 μ L binding buffer, 4 μ L Streptavidin beads, 15 μ L nuclease-free deionized water and mix carefully.
- (b) Dispense 55 μ L Binding Mix in each well of the 96-well PCR full skirted plate following the sequencing grid that was selected in **step 1**.
- (c) Add to these wells 25 μ L of biotinilated HLA amplification product.
- (d) Cover the plate with an adhesive transparent film (ClearSeal film for qPCR).
- (e) Mix on the microplate mixer at 1,200 rpm for at least 10 min. Be careful to maintain in agitation to avoid the sedimentation of Sepharorose beads.

3.3.4 Preparation of the PSQ 96 Plate Low Sequencing Plate

- (a) Prepare for each sample 40 μ L of annealing mix containing 33.6 μ L Annealing Buffer (*see Note 2*).
- (b) Add 6.4 μ L of 5 μ M HLA-C sequencing primer (HLA-C SEQ, Table 1) to a final concentration of 800 nM (*see Note 2*).
- (c) Dispense 40 μ L of Annealing Mix in each well of the sequencing grid in the “PSQ 96 Plate Low” carefully pipetting up and down repeatedly.

3.3.5 Single Stranded Template Purification

- (a) Recover the 96-well PCR full skirted plate after 10 min agitation and remove gently the adhesive film (to avoid possible cross-contaminations).
- (b) Turn on the vacuum pump and apply vacuum to the Prep Tool.
- (c) Wash the filter tips aspirating water for 20 s.
- (d) Capture the Sepharose beads by inserting the tool in the 96-well full skirted PCR plate with vacuum on. Make sure that all the liquid in the various wells has been aspirated and the beads are attached to the filter tips.
- (e) Move the PyroMark Vacuum-Prep Tool to the trough containing 70 % ethanol (position 2) and mix carefully for 5 s.
- (f) Move the tool to the Denaturation Solution trough (position 3) and wait 5 s while aspirating.
- (g) Gently move to the Washing buffer Trough (position 4) and leave it for 5 s.
- (h) Lift the PyroMark Vacuum-Prep Tool and place it for a few seconds in a vertical position, with an inclination greater than 90°, to allow the liquid to be completely aspirated.

	G	
C		T
S	A	E

Fig. 2 PSQ 96 reagent cartridge

- (i) Position the PyroMark Vacuum-Prep Tool above the “PSQ 96 Plate Low” and turn off the vacuum. To facilitate the release of the Sepharose beads gently move and shake the tool inside the wells, paying attention that the tool is always in the bottom of the wells.
- (j) Move the tool to the trough in position 5 that contains deionized water and shake for 5 s. Then apply the vacuum and wash the PyroMark Vacuum Prep Tool filters using the trough at the parking position folled with water for 20 s.
- (k) When filters have been cleaned turn off the vacuum.
- (l) Put the “PSQ 96 Plate Low” containing samples on the shaking incubator setting the instrument (*see Note 3*) Incubate for 2 min at 80 °C without shaking.

3.4 HLA-C Pyrosequencing

3.4.1 Pyrosequencing

- (a) Record the required volume of the E-mix, S-mix and nucleotides (A, C, G, T) available in the previously prepared working sheet page “SNP Run Set-up” (click view and select run).
- (b) Insert the PSQ cartridge and reconstitute (if necessary) mixtures of enzyme (E-) and substrate (S-) mix by adding 620 µL of sterile deionized water to the freeze-dried reagents. Leave at room-temperature for 5–10 min before dispensing into the PSQ 96 cartridge following the order shown in Fig. 2.
- (c) Set the nucleotide order of dispensation, insert the loaded cartridge into the instrument and start the Pyrosequencing run.

3.4.2 Pyrogram® Interpretation

Figure 3 shows an example of HLA-C supra-typing of samples belonging to C1/C1, C1/C2 or C2/C2 group of alleles following the indications given in Table 4.

See Note 4 for information on which HLA-C alleles are analyzed by the assay.

3.5 HLA-B Template Preparation

3.5.1 HLA-B Locus Specific PCR

- (a) For each sample, dispense 50 ng of DNA to each well of an eight-well microtube (0.2 mL) strip.
- (b) Add 21.5 µL of a PCR mix containing the reagent as described in Table 5 (*See Note 2*).
- (c) Adjust to 50 µL final volume with nuclease-free distilled/deionized water.

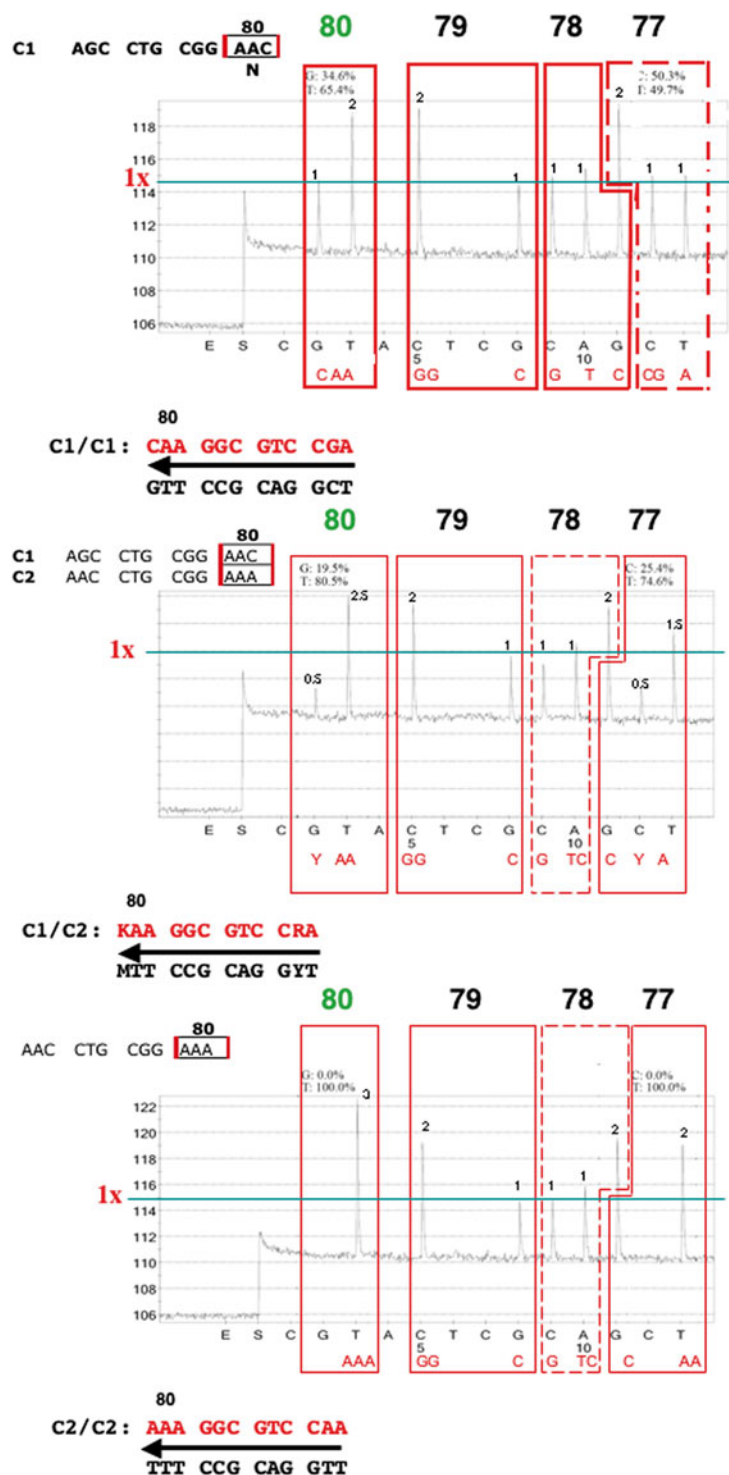


Fig. 3 Examples of pyrograms analyzing HLA-C (sample homozygous for C1/C1, heterozygous for C1/C2 and homozygous for C2/C2). Only the polymorphism in anticodon 80 is critical for the attribution to either the C1 or C2 group of alleles, anticodon 77 provides additional informative on the classification for the majority of the alleles

Table 4
HLA-C pyrogram interpretation

Typing result	Anti codon 80 [G/T] ^a	Anti codon 77 [C/T]	Alleles
C1/C1	G >30 % and T <70 %	C ~50 % and T ~50 %	The majority of C1
C1/C1	G >30 % and T <70 %		<i>See Note 4</i>
C1/C2	G <20 % and T >80 %	C <30 % and T >70 %	The majority of C1/C2
C1/C2	G <20 % and T >80 %		<i>See Note 4</i>
C2/C2	T 100 %	T 100 %	The majority of C2
C2/C2	T 100 %		<i>See Note 4</i>

^aThe SNP (anticodon 80) is the only critical residue for attribution to the C1/C2 group, while the anti codon 77 is to be considered only for the classification of the majority of the alleles (*see Note 4*)

Table 5
HLA-B locus specific PCR mix

Reagent	μL	[Concentration] _f
10× Buffer	5.00	
10 mM dNTP mix	1.00	0.2 mM
50 mM MgCl ₂	1.50	1.5 mM
Platinum Taq (5 U/μL)	0.50	2.5 U
HLA-B BEX2FRW (10 μM)	7.5	1,500 nM
HLA-B BEX2REVb (10 μM)	2	400 nM
5Bln1-57 (10 μM) [19]	2	400 nM
3Bln3-37 (10 μM) [19]	2	400 nM
Total volume mix	21.5	

- (d) Mix carefully.
- (e) Cap the strip and start PCR cycling as described in **step (d)** of Subheading 3.2.1.

3.5.2 HLA-B Nested PCR

- (a) Prepare the HLA-B nested reagent mix (*see* Tables 1 and 6) and dispense 24 μL of this mix for each sample in a new eight-well microtube (0.2 mL) strip (*See Note 2*).
- (b) Add 1 μL of the first HLA-B locus specific PCR product to each well of a the eight-well microtube (0.2 mL) strip,
- (c) Cap the strip and start PCR cycling as described in **step (c)** of Subheading 3.2.2.

Table 6
HLA-B nested PCR mix

Reagent	μL	[Concentration] _f
10× Buffer	2.50	
10 mM dNTP mix	0.50	0.2 mM
50 mM MgCl ₂	0.75	1.5 mM
Platinum Taq (5 U/ μL)	0.25	1.25 U
Biot-HLA-B Frw (10 μM)	0.25	100 nM
HLA-B Rev (10 μM)	0.5	200 nM
H ₂ O	19.25	
Total volume mix	24	

3.6 Setting Up the Pyrosequencer

Follow the Subheadings 3.3.1–3.3.5 and modify in Subheadings 3.3.1 and 3.3.4 the following points:

- Set the ID ENTRY to HLA-B supratyping.
- Enter the nucleotide dispensation order: TAC gTC AgC ATg TCg ATC gCA TCg Cag.
- Enter the sequence to be analyzed: ACg TCA g[C/A]g TC[g/A] CgC ATC gCA gCC gCA g
- Use the HLA-B specific sequencing primer HLA-B SEQ in **step b** of Subheadings 3.3.4 (Table 1).

3.7 HLA-B Pyrogram Interpretation

Examples of HLA-B supra-typing are given in Fig. 4a, b with the different combinations of Bw6/Bw6, Bw6/Bw4 T₈₀A₈₁, Bw6/Bw4 T₈₀L₈₁, Bw6/Bw4 I₈₀, Bw4 T₈₀A₈₁/Bw4 T₈₀A₈₁, Bw4 T₈₀L₈₁/Bw4 T₈₀L₈₁, Bw4 I₈₀/Bw4 I₈₀, Bw4 T₈₀A₈₁/Bw4 T₈₀L₈₁, Bw4 T₈₀A₈₁/Bw4 I₈₀ or Bw4 T₈₀L₈₁/Bw4 I₈₀ group of alleles following the indication for the analysis described in Table 7. See **Note 5** for information on the analyzed HLA-B alleles.

3.8 HLA-B*46 Allotypes Characterized by HLA-C C1-Group Specificity

3.8.1 B*46 Specific Template Preparation

- Use primers: HLA-B 5Bin1-57 and HLA-C 3CIn3-12 both at 400 nM (Table 1).
- Amplify by PCR using the following protocol: 10 min denaturation at 95 °C, 35 cycles of 30 s denaturation at 95 °C, 1 min at 62 °C for primer annealing and 30 s elongation at 72 °C. The program is terminated with a final extension step for 5 min at 72 °C.

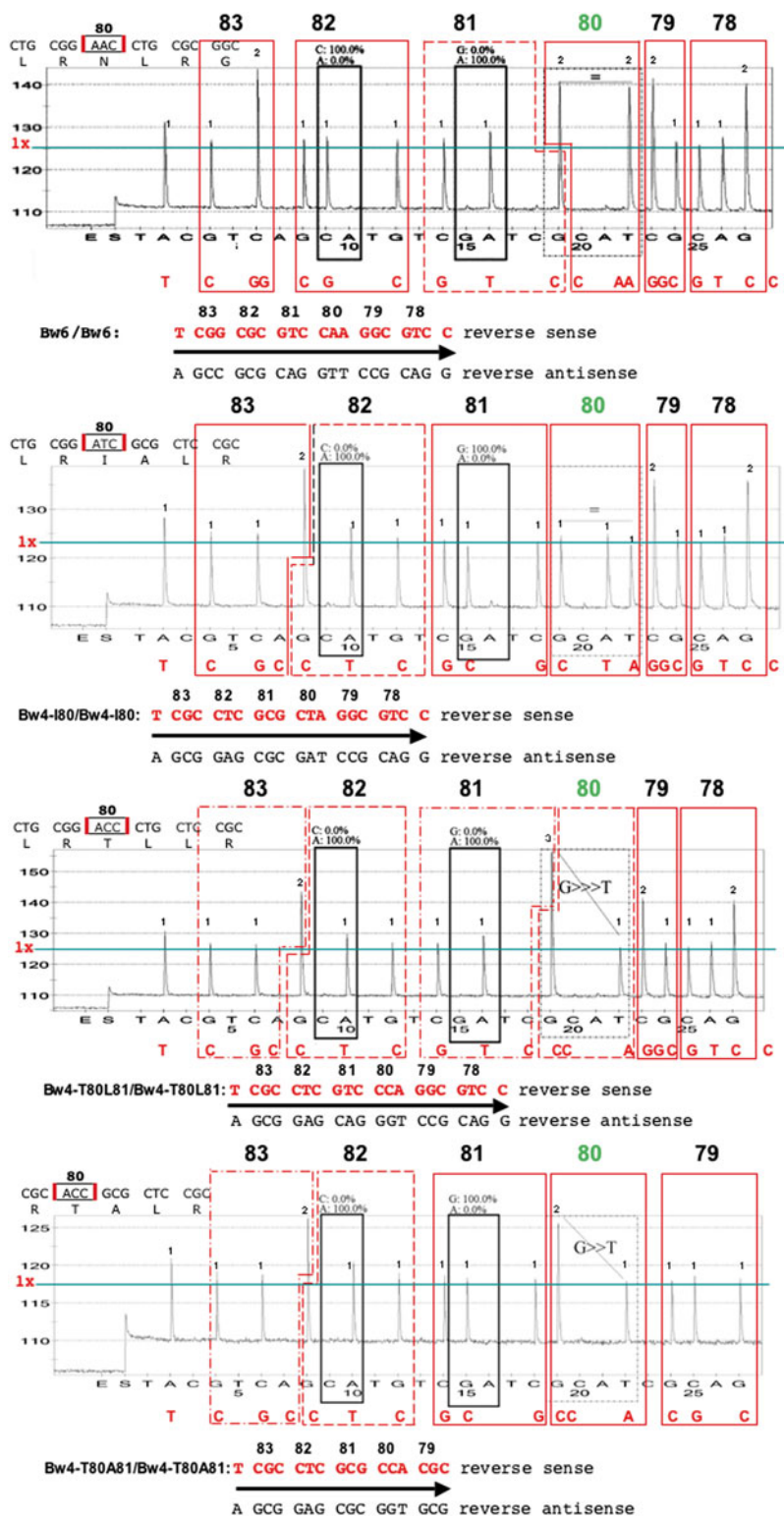


Fig. 4 (a) Examples of Pyrograms analyzing HLA-B for homozygous samples (Bw6/Bw6, Bw4-I⁸⁰/Bw4-I⁸⁰, Bw4-T⁸⁰L⁸¹/Bw4-T⁸⁰L⁸¹, Bw4-T⁸⁰A⁸¹/Bw4-T⁸⁰A⁸¹). (b) Examples of Pyrograms analyzing HLA-B for heterozygous samples (Bw6/Bw4-I⁸⁰, Bw6/Bw4-T⁸⁰L⁸¹, Bw6/Bw4-T⁸⁰A⁸¹, Bw4-I⁸⁰/Bw4-T⁸⁰A⁸¹)

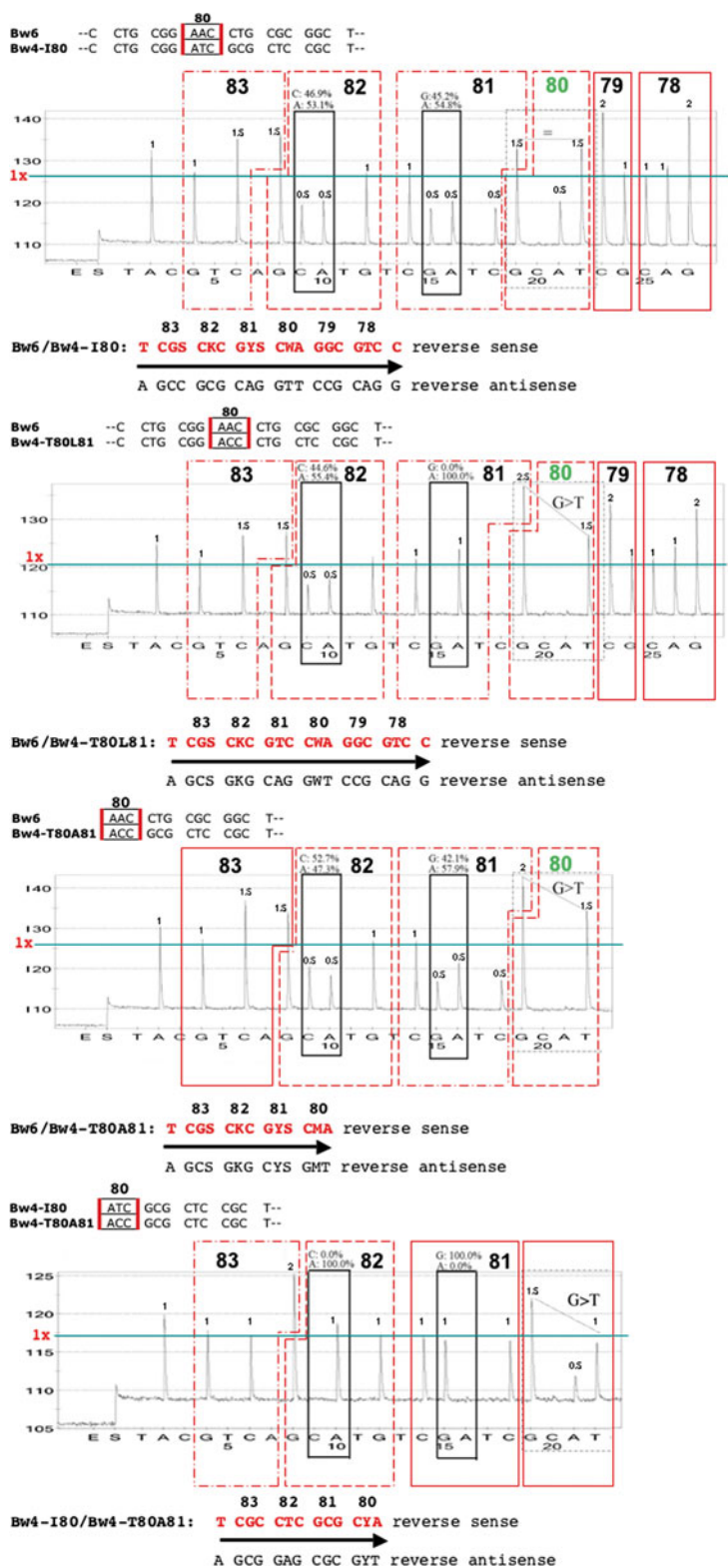


Fig. 4 (continued)

Table 7
HLA-C pyrogram interpretation

Typing	Anti codon 82 [C/A]	Anti codon 81 [G/A]	Anti codon 80 GCAT	Observed peak height
Bw6/Bw6	C	A	No A	G double height = T double height
Bw6/Bw4 T ₈₀ A ₈₁	C + A	G + A	No A	
Bw6/Bw4 T ₈₀ L ₈₁	C + A	A	No A	
Bw6/Bw4 I ₈₀	C + A	G + A	Half A peak	
Bw4 T ₈₀ A ₈₁ /Bw4 T ₈₀ A ₈₁	A	G	No A	G double height, T
Bw4 T ₈₀ L ₈₁ /Bw4 T ₈₀ L ₈₁	A	A	No A	G triple height, T
Bw4 I ₈₀ /Bw4 I ₈₀	A	G	A	G = A = T
Bw4 T ₈₀ A ₈₁ /Bw4 T ₈₀ L ₈₁	A	G + A	No A	
Bw4 T ₈₀ A ₈₁ /Bw4 I ₈₀	A	G	Half A peak	G > T > A
Bw4 T ₈₀ L ₈₁ /Bw4 I ₈₀	A	G + A	Half A peak	

3.8.2 B*46 Nested Amplification

- Use primers: biot-HLA-B B46 Frw and HLA-B B46 Rev (Table 8).
- Amplify by PCR using the following protocol: 2 min denaturation at 95 °C, 40 cycles of 15 s denaturation at 95 °C, 15 s at 60 °C for primer annealing, 30 s at 72 °C. A single elongation step of 5 min at 72 °C conclude the amplification.

3.8.3 Follow the Subheadings 3.3.1–3.3.5 and Modify in Subheadings 3.3.1 and 3.3.4 the Following Points

- Set the ID ENTRY to HLA-B46 supratyping.
- Enter the nucleotide dispensation order: CTCATgCTCgAgTCgAgTCACgTgT.
- Enter the sequence to be analyzed: TC[A/T]CTCgAgCTCgAg[T/C]CgTgT.
- Use the HLA-B46 SEQ (specific sequencing primer) in **step b** of Subheadings 3.3.4 5'-gCA ggT TCC gCA ggC-3' at 800 nM.
- For the interpretation of the Pyrograms (Fig. 5a) *see* Table 9 and **Note 6** for information on the analyzed HLA-B46 alleles.

3.9 HLA-B*73 Allotypes Characterized by HLA-C C1-Group Specificity**3.9.1 B*73 Specific Template Preparation**

- Use the primers HLA-B 5Bin1-57 and HLA-C 3CIn3-12 both at 400 nM (*see* Table 1).
- Amplify by PCR using the following protocol: 10 min denaturation at 95 °C, 35 cycles of 30 s denaturation at 95 °C, 1 min at 62 °C for primer annealing and 30 s elongation at 72 °C. The program is terminated with a final extension step for 5 min at 72 °C.

Table 8
HLA-B*46 and HLA-B*73 primers

Allotype	Primer name	Primer sequence	Concentration (nM)
HLA-B*46	5Bln1-57 [19]	GGG AGG AGC GAG GGG ACC RCA G	400
	3Cln3-12 [19]	GGA GAT GGG GAA GGC TCC CCA CT	400
	Biot-HLA-B B46 Frw	Biotin-GGA CCG GGA GAC ACA GAA G	400
	HLA-B B46 Rev	GGC CTC GCT CTG GTT GTA G	400
	HLA-B B46 SEQ	GCA GGT TCC GCA GGC	800
HLA-B*73	5Bln1-57 [19]	GGG AGG AGC GAG GGG ACC RCA G	400
	3Cln3-12 [19]	GGA GAT GGG GAA GGC TCC CCA CT	400
	HLA-B B73 Frw90	ACA GAT CTG CAA GGC CAA G	900
	Biot-HLA-B B73 Rev90	Biotin-GCC ATA CAT CGT CTG CCA A	100
	HLA-B B73 SEQ	GGC ACA GAC TGA CCG	800

3.9.2 B*73 Nested Amplification

- Use primers: HLA-B B73 F90 and HLA-B B73 biotin-R90 (Table 8).
- Amplify by PCR using the following protocol: 2 min denaturation at 95 °C, 40 cycles of 15 s denaturation at 95 °C, 15 s at 60 °C for primer annealing and and 30 s elongation at 72 °C. The program is terminated with a final extension step for 5 min at 72 °C.

3.9.3 Follow the Subheadings 3.3.1–3.3.5 and Modify in Subheadings 3.3.1 and 3.3.4 the Following Points

- Set the ID ENTRY to HLA-B73 supratyping.
- Enter the nucleotide dispensation order: CAg TAT gAg ACT g.
- Enter the sequence to be analyzed: Ag[T/A] g[A/g]g CCT gCg gAA CCT gCg Cgg CTA C.
- Use the HLA-B73 SEQ sequencing primer 5'-ggC ACA gAC TgA CCg-3' at 800 nM.
- Interpret Pyrograms (Fig. 5b) following the instructions given in Table 10.

4 Notes

- Always wear disposable powder-free gloves in each area and change them quite often during the procedure to decrease the possibility of sample contamination.
- Calculate the required volume for the Mix by multiplying the volume needed for a single sample with the number of samples to be analyzed adding at least two additional volumes every 48 samples.
- This step is required to anneal the short sequencing primer to the single stranded template.

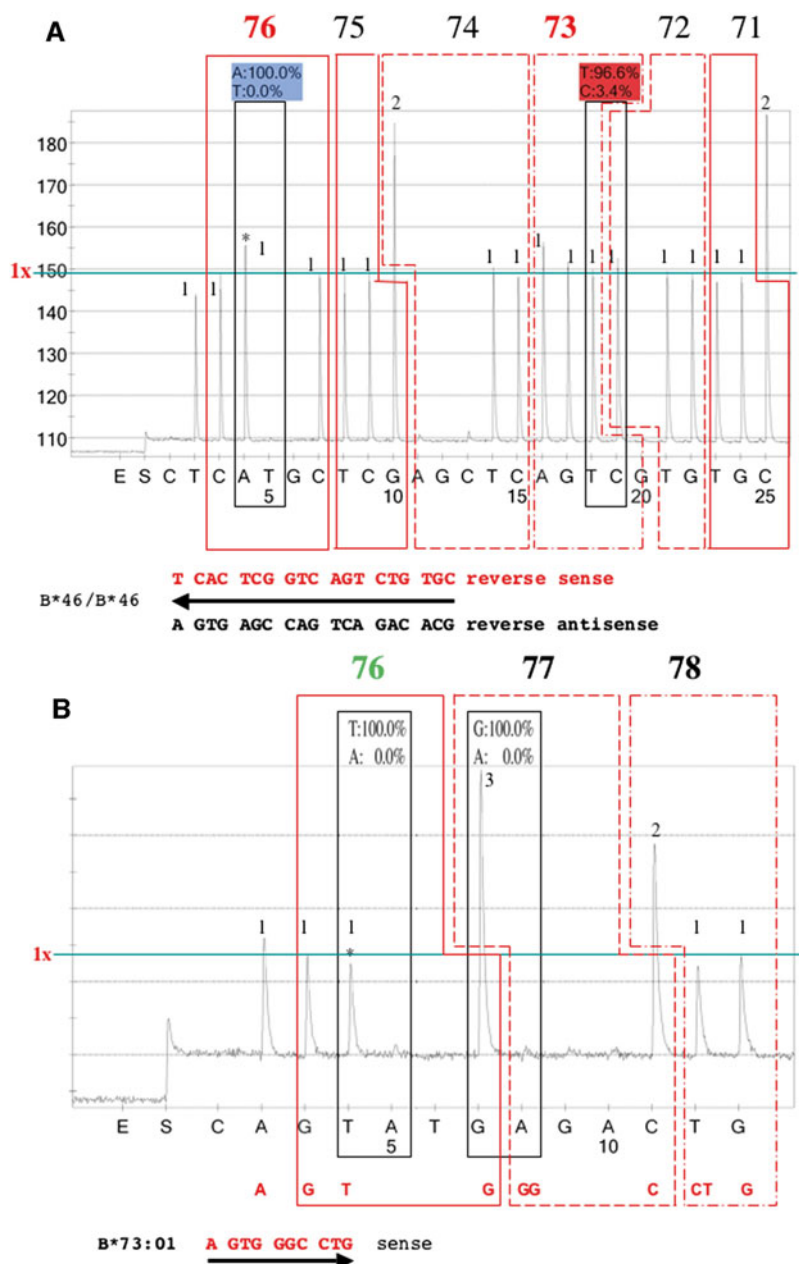


Fig. 5 Examples of Pyrograms for HLA-B alleles that share sequences of the HLA-C1 group. (a) Sample for cell line IHW: 9126, WATANABE (HLA-B*46 homozygous sample). *critical residue identifying all HLA-B*46 alleles and the contaminant coamplified B*07:13 allele, the latter is recognized and discarded analyzing the last base of codon 73 (second polymorphism analyzed that need to be for B*46T (and not C) in the pyrogram. (b) The sample obtained from the IHW: 9280, LK707 cell line is known to be HLAB*73:01/HLA-B*52:01:01. Since the B*52 allele is not amplified by the specific primer combination used, the sample is to be considered as a B*73 homozygous sample. * critical residue identifying all HLA-B*73 alleles

Table 9
HLA-B*46 pyrogram interpretation

Typing	Anti codon 76 [T/A]	Anti codon 73 [T/C]	Additional bases at dispensation
B*46:01:01 ^a	A	T	
B*46:01:02	A	T	A-11
B*46:02	A	T	C-14 (double peak height)
B*46:35	A	T	G-15
B*46:50	A	T	A-16 (double peak height)
B*46:37	A	T	C-22
B*07:13	A	C (double peak height)	
B*67:02	A	C (double peak height)	

^aSee Note 6

Table 10
HLA-B*73 pyrogram interpretation

Typing	Codon 76 [T/A]	Codon 77 [G/A]
B*73:01	T	G
B*73:02	T	G

- Analyzing all the 2133 sequences of the HLA-C alleles present in the IMGT database in March 2014, (<http://www.ebi.ac.uk/ipd/imgt/hla/align.html>) the described procedure will miss six alleles (C*03:33, C*03:46, C*03:63, C*07:80, C*07:201, and C*08:81) accounting for 0.28 % of all the known HLA-C alleles. In detail, 1318 alleles belong to the group C1 (Asn80), between them the majority are Ser77/Asn80, but 14 are characterized by the presence of Asn77/Asn80 (C*01:59, C*02:65, C*03:130, C*03:140, C*04:114, C*05:20, C*06:82, C*07:49, C*07:210, C*07:238, C*07:247, C*12:54, C*14:04, C*16:57). On the contrary, 811 belong to the group C2 (Lys80) with the majority of C2 alleles being Asn70/Lys80, but some of them are characterized by Ser77/Lys80 (C*02:12, C*03:10, C*03:29, C*06:122, C*12:72, C*15:21, and C*16:37).
- Analyzing all the 3285 sequences of HLA-B alleles present in the IMGT database, in March 2014, (<http://www.ebi.ac.uk/ipd/imgt/hla/align.html>) the described procedure will miss 15 alleles, four being Null alleles (B*08:92, B*13:28, B*15:79 N, B*15:158, B*18:84, B*35:221, B*40:01:35, B*40:02:15, B*40:22 N, B*40:142 N, B*40:168, B*40:175, B*56:28 N, B*58:29, B*58:42) accounting for 0,46 % of all known HLA-B alleles.

6. Identical results will be obtained for the following alleles:
 B*46:01:03, B*46:01:04, B*46:01:05, B*46:01:06, B*46:01:07,
 B*46:01:08, B*46:01:09, B*46:01:10, B*46:01:11,
 B*46:01:12, B*46:01:13, B*46:01:14, B*46:01:15, B*46:03,
 B*46:04, B*46:05, B*46:06, B*46:07 N, B*46:08, B*46:09,
 B*46:10, B*46:11, B*46:12, B*46:13:01, B*46:13:02,
 B*46:13:03, B*46:14, B*46:15 N, B*46:16, B*46:17,
 B*46:18, B*46:19, B*46:20, B*46:21:01, B*46:21:02,
 B*46:22, B*46:23, B*46:24, B*46:25, B*46:26, B*46:27,
 B*46:28, B*46:29, B*46:30, B*46:31, B*46:32, B*46:33,
 B*46:34, B*46:36, B*46:37, B*46:38, B*46:39, B*46:40,
 B*46:41 N, B*46:42, B*46:43, B*46:44, B*46:45, B*46:46,
 B*46:47, B*46:48, B*46:49, B*46:50, B*46:51Q, B*46: 52,
 B*46:53, B*46:54, B*46:55 N, B*46:56, B*46:57.

References

1. Biassoni R (2009) Human natural killer receptors and their ligands. *Curr Protoc Immunol* 14:10
2. Ciccone E, Viale O, Pende D et al (1988) Specific lysis of allogeneic cells after activation of CD3-lymphocytes in mixed lymphocyte culture. *J Exp Med* 168:2403–2408
3. Kärre K (2008) Natural killer cell recognition of missing self. *Nat Immunol* 9:477–480
4. Parham P, Norman PJ, Abi-Rached L et al (2012) Human-specific evolution of killer cell immunoglobulin-like receptor recognition of major histocompatibility complex class I molecules. *Philos Trans R Soc Lond B Biol Sci* 367:800–811
5. Ljunggren HG, Kärre K (1985) Host resistance directed selectively against H-2 deficient lymphoma variants. Analysis of the mechanism. *J Exp Med* 162:1745–1759
6. Biassoni R, Falco M, Cambiaggi A et al (1995) Single aminoacidic substitutions can influence the NK-mediated recognition of HLA-C molecules. Role of serine-77 and lysine-80 in the target cell protection from lysis mediated by “group 2” or “group 1” NK clones. *J Exp Med* 182:605–609
7. Mandelboim O, Reyburn HT, Vales-Gomez M et al (1996) Protection from lysis by natural killer cells of group 1 and 2 specificity is mediated by residue 80 in human histocompatibility leukocyte antigen C alleles and also occurs with empty major histocompatibility complex molecules. *J Exp Med* 184:913–922
8. Winter CC, Gumperz JE, Parham P et al (1998) Direct binding and functional transfer of NK cell inhibitory receptors reveal novel patterns of HLA-C allotype recognition. *J Immunol* 161:571–577
9. Gumperz JE, Barber LD, Valiante NM (1997) Conserved and variable residues within the Bw4 motif of HLA-B make separable contributions to recognition by the NKB1 killer cell-inhibitory receptor. *J Immunol* 158: 5237–5241
10. Cella M, Longo A, Ferrara GB et al (1994) NK3-specific natural killer cells are selectively inhibited by Bw4-positive HLA alleles with isoleucine 80. *J Exp Med* 180:1235–1242
11. Luque I, Solana R, Galiani MD et al (1996) Threonine 80 on HLA-B27 confers protection against lysis by a group of natural killer clones. *Eur J Immunol* 26:1974–1977
12. Moesta AK, Abi-Rached L, Norman PJ et al (2009) Chimpanzees use more varied receptors and ligands than humans for inhibitory killer cell Ig-like receptor recognition of the MHC-C1 and MHC-C2 epitopes. *J Immunol* 182:3628–3637
13. Abi-Rached L, Moesta AK, Rajalingam R et al (2010) Human specific evolution and adaptation led to major qualitative differences in the variable receptors of human and chimpanzee natural killer cells. *PLoS Genet* 6:e1001192
14. Moesta AK, Norman PJ, Yawata M et al (2008) Synergistic polymorphism at two positions distal to the ligand-binding site makes KIR2DL2 a stronger receptor for HLA-C than KIR2DL3. *J Immunol* 180:3969–3979
15. Ugolotti E, Vanni I, Raso A et al (2011) Human leukocyte antigen-B (-Bw6/-Bw4 180, T80) and human leukocyte antigen-C (-C1/-C2)

- subgrouping using pyrosequence analysis. *Hum Immunol* 72:859–868
16. Biassoni R, Vanni I, Raso A et al. An improvement method for HLA-B C supratyping. Submitted for publication
 17. Boom R, Sol CJ, Salimans MM et al (1990) Rapid and simple method for purification of nucleic acids. *J Clin Microbiol* 28:495–503
 18. Erlich RL, Jia X, Anderson S et al (2011) Next generation sequencing for HLA typing of class I loci. *BMC Genomics* 12:42
 19. Cereb N, Maye P, Lee S et al (1995) Locus-specific amplification of HLA class I genes from genomic DNA: locus-specific sequences in the first and third introns of HLA-A, -B, and -C alleles. *Tissue Antigens* 45:1–11

Free ebooks ==> www.Ebook777.com

Chapter 12

Allele Quantification Pyrosequencing® at Designated SNP Sites to Detect Allelic Expression Imbalance and Loss-of-Heterozygosity

Chau-To Kwok and Megan P. Hitchins

Abstract

Pyrosequencing® is able to quantitate the level of a nucleotide at a designated germ-line or somatic variant, including single nucleotide polymorphisms (SNPs). SNPs within a gene of interest may be used to distinguish between the two genetic alleles and study their behavior in heterozygous individuals. With regard to cancer etiology and development, identification of alleles and the detection of allelic imbalances, such as transcriptional loss from one allele or loss-of-heterozygosity (due to deletion of one allele), within a tumor are particularly useful. Lynch syndrome, the most common form of hereditary bowel and uterine cancer, is caused by heterozygous germ-line mutations within the DNA mismatch repair genes and tumors develop following inactivation of the remaining functional allele within somatic tissues, usually by acquired loss-of-heterozygosity. *MLH1* is the most frequently mutated gene in Lynch syndrome; however, some cases whose tumors display immunohistochemical loss of the MLH1 protein have no apparent mutation within the coding region of *MLH1*. Allelic loss of expression or reduced function of *MLH1* can also result in the propensity to develop Lynch syndrome associated cancers. In this chapter we describe allele quantification Pyrosequencing assays designed at a common benign SNP within the *MLH1* coding region for application to either DNA or mRNA (cDNA) templates, which enabled us to detect pathological allelic imbalances in such cases with suspected Lynch syndrome. Our allele quantification Pyrosequencing assays at the *MLH1* c.655A>G (rs1799977) exonic SNP were applied to clinical specimens and detected both constitutional allelic expression loss and tumor loss-of-heterozygosity in some cases, facilitating the identification of the mechanistic cause underlying their cancer development. We provide detailed protocols for implementing these Pyrosequencing assays and illustrative examples of their application in patients.

Key words Pyrosequencing®, Allele quantification, Genotyping, Allelic expression, Loss of heterozygosity

1 Introduction

Pyrosequencing® is a real-time quantitative bioluminescence technique used for DNA sequencing that can measure the relative levels of a given nucleotide at a predetermined sequence. This technique uses an enzymatic reaction whereby, as each

nucleotide is incorporated during DNA elongation, a pyrophosphate is released. This pyrophosphate is then converted into ATP and subsequently used by luciferase to generate light. The amount of light ultimately emitted is proportional to the amount of incorporation of a particular nucleotide at a designated position within the DNA chain [1, 2]. This technique thus lends itself to accurate allele quantification, that is, the measurement of the relative levels of two genetic alleles, at a specified location within a DNA sequence. In the context of cancer etiology, allele quantification Pyrosequencing has been applied to the detection of specific germ-line mutations that occur frequently in particular populations and predispose to cancer development, including common mutations within the *MYH* gene associated with colorectal polypsis and cancer [3]. We and others have also used this method to detect recurrent somatic mutations in carcinoma, including the oncogenic *BRAFV600E* and *KRAS*, codon 12 and 13 mutations respectively, within colorectal carcinoma, which are useful in directing treatment options [4].

Pyrosequencing can also be adapted to the investigation of gene expression. One particular circumstance where Pyrosequencing is invaluable is in the measurement of the relative levels of transcription from individual alleles of a gene at a designated heterozygous locus within the gene of interest. In individuals who are heterozygous for an exonic genetic variant, be it a SNP or a suspected germ-line mutation, this locus can be exploited to differentiate between the two alleles and measure the relative levels of the transcripts derived from each allele. If the gene is expressed equally (biallelically) from both alleles, equivalent levels of each allele of the tracer SNP will be detected among the mRNA transcripts. This is referred to herein as “balanced allelic expression.” When the expression of one allele is altered due to a pathological defect, such as the presence of aberrant promoter methylation or a frameshift mutation that results in nonsense-mediated mRNA decay, then the level of one allele will exceed the other at the tracer SNP within the corresponding mRNA. This is referred to generally as an “allelic expression imbalance.” In some cases, although two alleles are detected at equal levels within the genomic DNA of heterozygous individuals, only one allele is detected among the mRNA transcripts, due to monoallelic expression. This strategy has been used to determine if genes of interest are subject to genomic imprinting, whereby only one parental allele is expressed, and the other is naturally inactivated within somatic tissues [5, 6], as well as disruptions in normal genomic imprinting expression patterns in disease settings (reviewed in [7]).

We adopted the principle of allele quantification by Pyrosequencing to study the allelic behavior of a non-imprinted

cancer-related gene to identify pathological changes in gene expression, specifically loss or significant reduction of expression of one allele within normal tissues that may have predisposed to cancer development. Furthermore, by comparing the allele frequency in genomic DNA of normal versus diseased tissue, one can determine the presence of “loss-of-heterozygosity” using allele quantification assays. Loss-of-heterozygosity as a consequence of genetic deletion or chromosomal loss is a frequent occurrence in cancer. This can result in diminished expression of protective genes, or can serve as the “second hit” in the somatic tissues of individuals predisposed to cancer development due carriage of an autosomal dominant heterozygous germ-line mutation, resulting in the initiation of tumorigenesis [8].

This chapter focuses on the application of allele quantification Pyrosequencing assays designed at a specific target SNP within the DNA mismatch repair gene, *MLH1*, which is frequently mutated in the autosomal dominant cancer condition Lynch syndrome. In Lynch syndrome, the initial defect within the DNA mismatch repair gene may be a point mutation, a large structural rearrangement such as a deletion, or constitutional methylation of the *MLH1* promoter (referred to as a constitutional epimutation). Current genetic screening practices involve screening for point mutations within the coding exons and for large genetic rearrangements. However, a significant proportion of Lynch syndrome cases whose tumors demonstrate *MLH1* protein loss, thereby implicating a germ-line defect within the *MLH1* gene, have no apparent germ-line mutation. Some of these cases may carry a cryptic mutation such as an intronic frameshift mutation (by creating a donor splice site), promoter mutation or constitutional epimutation, which can give rise to an allelic expression imbalance. Our aim was to develop allele quantification Pyrosequencing assays at a common SNP site within an *MLH1* exon, which could be exploited to identify allelic expression imbalances in mRNA, as well as allelic genetic losses in tumor DNA, in such patients with probable Lynch syndrome. The c.655A>G (rs1799977) SNP was selected for this purpose because it is abundant within the population and is exonic, located within exon 8, and therefore present within the mRNA transcript. Separate allele quantification Pyrosequencing assays were designed for genomic DNA templates and mRNA templates at the same c.655A>G SNP site and applied to the clinical specimens of suspected Lynch syndrome patients with *MLH1*-deficient tumors. These assays provided complementary testing to identify allelic genetic losses within tumor tissues and allelic expression imbalances within normal tissues, respectively [9].

2 Materials

2.1 PCR and Pyrosequencing® Primers and Reagents for Allele Quantification of the *MLH1* c.655A > G (rs1799977) SNP

1. PCR reagents.
2. Invitrogen Platinum® Taq.
3. 10× PCR Buffer, without Magnesium chloride.
4. 50 mM magnesium chloride.
5. 10 mM dNTP (deoxynucleotide triphosphates) mixture.
6. 10 mM of each primer (only one biotinlyated). The primer sequence and the PCR conditions for interrogating the rs1799977 SNP in *MLH1* are listed in Table 1.

2.2 Agarose Gel Electrophoresis

1. Agarose A.
2. 0.5× TBE (Tris–Borate–EDTA) buffer: 445 mM Tris, 445 mM boric acid, 10 mM EDTA in H₂O.
3. Gel loading dye.
4. Nucleic acid stain (for example, gel red).

2.3 Pyrosequencing Reagents and Consumables

1. PyroMark® ID using PyroMark ID software version 1.0.
2. PSQ 96 plate low.
3. Streptavidin Sepharose® High Performance beads.
4. PyroMark Q96 Reagent Cartridge.
5. PyroMark Gold Q96 Reagents.
6. Pyrosequencing troughs
7. Binding buffer: 10 mM Tris–HCl, 2 M NaCl, 1 mM EDTA, 0.1 % Tween™ 20, pH 7.6.
8. Annealing buffer pH 7.6: 20 mM Tris–HAcetate, 2 mM Mg–Acetate.

Table 1
Pyrosequencing oligonucleotide sequences and conditions

Gene site and primer use Oligonucleotide sequences		Conditions		
		Size (bp)	[MgCl ₂] (mM)	T _m (°C)
<i>MLH1</i> c.655A > G (rs1799977)				
gDNA PCR	F: 5′-GCCTCAACCGTGGACAATATTC-3′ R: biotin-5′-CGACATACCGACTAACAGCATTTTC-3′	58	2	56
cDNA PCR	F: 5′-GCCTCAACCGTGGACAATATTC-3′ R: biotin-5′-GCTACGGTTTTATCCTCACATCCA-3′	89	2	64
Pyrosequencing	5′-GGACAATATTCGCTCC-3′	NA		

9. 70 % ethanol.
10. Denaturing buffer: 0.2 M NaOH.
11. Washing buffer: 10 mM Tris–HAcetate, pH 7.6.
12. 4 M acetic acid for pH adjustment.
13. 1 M HCl for pH adjustment.

3 Methods

3.1 Outline of the Pyrosequencing Protocol

An outline of the steps involved in the pyrosequencing process for allele quantification is shown in Fig. 1. This is followed by a more detailed protocol.

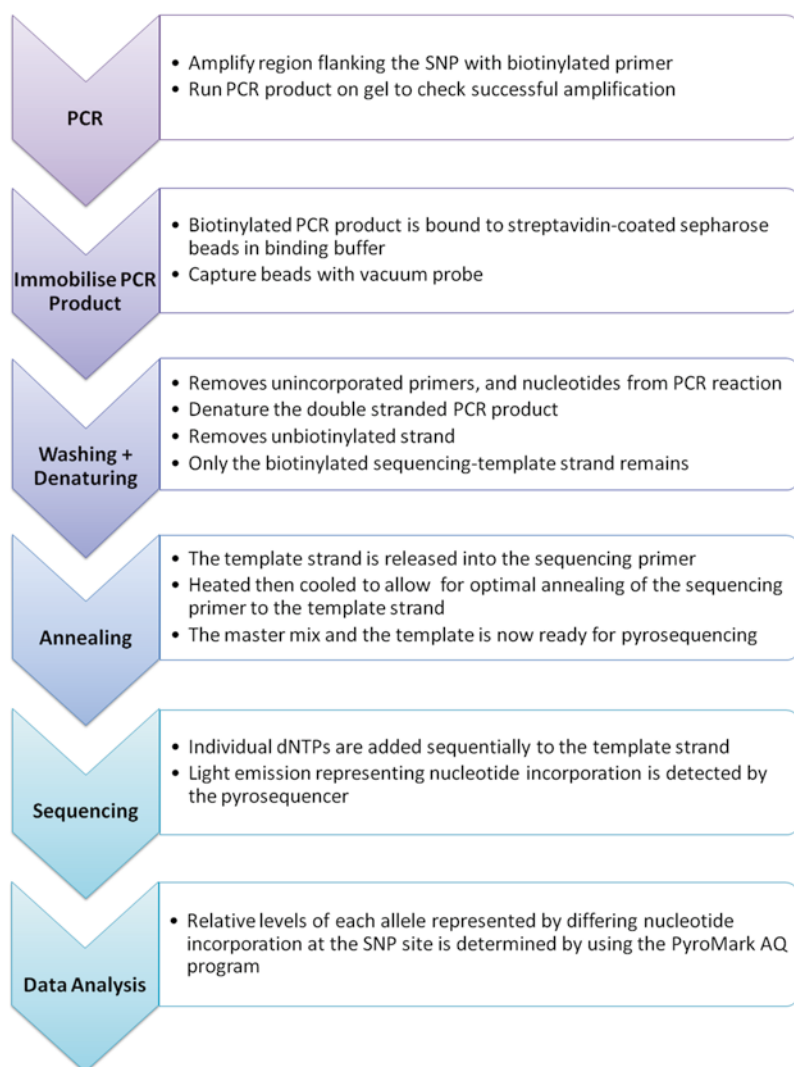


Fig. 1 Pyrosequencing flow diagram. A step-by-step summary of the procedures involved in Pyrosequencing

3.2 Assay Design for Measuring the Relative Levels of Each Allele at the *MLH1* c655A>G SNP (rs1799977) Within Genomic DNA and cDNA

An allele quantification assay was designed for measuring the relative levels of each of the two genetic alleles at a particular SNP site, c.655A>G (rs1799977), within exon 8 of the *MLH1* gene for both genomic DNA and cDNA templates. The assay designed for genomic DNA was applied to detect loss-of-heterozygosity within tumor DNA, as compared to DNA from normal tissue. The cDNA assay was designed to detect transcriptional loss or reduction from one allele of *MLH1* compared to the other. In either case, the SNP was used as the tracer of the two distinct genetic alleles of *MLH1* in subjects heterozygous for this common SNP. Two notable points in the design of this assay, which are also applicable for the design of any assay for measuring relative levels of allelic expression of a gene are: (1) PCR amplification primers for the cDNA template should ideally span at least one intron, i.e., cross two exons (one of which contains the target SNP site). This allows for easy elimination or detection of genomic DNA contamination post-PCR. In this case, the assay was designed with the forward RT-PCR primer within *MLH1* exon 8, which also contained the SNP site, and the reverse primer, which was located within exon 9. (2) The PCR products from DNA and cDNA templates should be similar in size and the same nested Pyrosequencing primer and nucleotide dispensation order should be used to assay the target SNP, since this will minimize any systematic bias between the assays from the two sources of template. This is important because the peak heights detected for each allele within the cDNAs will be normalized against the peak heights of the same respective allele in the genomic DNA. In this case, the same forward primer and Pyrosequencing primer were used for both genomic DNA and cDNA templates, but the reverse primer differed (for genomic DNA this was located partially within intron 8, whereas for cDNAs this was necessarily located within exon 9).

The assay maps for allele quantification Pyrosequencing of the *MLH1* c.655A>G SNP within genomic DNA and cDNA templates are shown in Fig. 2. The assay was designed for use on the PyroMark Q96 ID pyrosequencing platform using the accompanying software, PyroMark IQ and Pyrosequencing (PSQ) Assay Design Software Version 1.0.

3.3 Nucleic Acid Preparation

DNA and RNA extractions can be carried out using standard laboratory techniques. The cDNA may be prepared from 500 to 1,000 ng of total RNA and converted to cDNAs using any reverse transcriptase kit. The Invitrogen Superscript III was used in this case, as per the manufacturer's instruction.

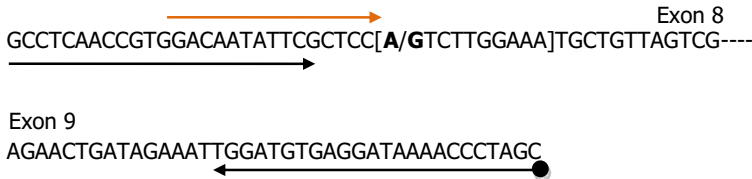
3.4 PCR

PCR is then used to amplify the region of interest using approximately 20 ng of DNA or cDNA (equivalent of about 50–100 ng of RNA). The PCR primer for the template strand to be sequenced (when subsequently denatured yielding a single strand) was

Genomic DNA



cDNA



SNP and usage	Dispensation Order	Sequence examined
Rs1799977	C-A-G-A-T-C-T-G-A	5'-A/GTCTTTGAAAA-3'.
<i>MLH1</i> c.655A>G		

Fig. 2 Map of Pyrosequencing assays at *MLH1* c.655A>G (rs1799977). Maps of the PCR amplicons for the Pyrosequencing assays at the c.655A>G SNP within exon 8 of *MLH1* for both genomic DNA (*top*) and cDNA (*bottom*) are shown. Exonic sequence is in upper case and intronic sequence is in lower case. The hyphens in the cDNA sequence show the exon boundary. The SNP is in *bold*, and sequences within the brackets are the bases to be examined in the Pyrosequencing reaction. The *arrows* indicate the positions of the primers; *black* indicates the PCR primers, orange the nested Pyrosequencing primer. The *black circle* represents the biotinylation of the 5' end of the reverse PCR primers for the subsequent isolation of this strand as the Pyrosequencing template. The table illustrates the SNP investigated, the nucleotide dispensation order of the pyrosequencer and the *MLH1* sequence interrogated

5'-biotinylated, and was HPLC-purified. Conditions for individual PCR primer sets should be optimized to produce a single and specific amplification product. The PCR set-up and cycling conditions for this assay were as follows:

1. Thaw all reagents at room temperature except for the DNA polymerase, which is kept on ice.
2. Mix all reagents thoroughly to equilibrate ionic concentrations (do not vortex the DNA polymerase), and centrifuge briefly to collect aerosols.
3. Prepare a master mix for all the samples including a no-template control.
4. Ensure that the master mix is mixed thoroughly and centrifuge briefly. Aliquot 28 µL of the master mix into each tube. Add 2 µL of template into each tube, except the no-template control.
5. Pulse-centrifuge the PCR tubes (to remove any air bubbles and collect the mix at the bottom) and place them into the PCR machine and run the appropriate conditions for each PCR.
6. PCR conditions for the amplification of genomic DNA at rs1799977 are: 95 °C-5 min initial denaturation followed by up to 40 cycles of 95 °C-30 s, 56 °C-30 s, 72 °C-30 s.

PCR conditions for amplification of cDNAs at rs1799977 are: 95 °C—5 min; up to 40 cycles of 95 °C—30 s, 64 °C—30 s, 72 °C—30 s.

7. Run 5 µL of the PCR amplified product on a 2 % agarose gel to confirm PCR amplification products are visible and there has been no contamination of the negative (no sample added) control.
8. PCR products may be stored frozen at −20 °C for Pyrosequencing at a future date. No purification of the PCR products is required.

3.5 Pyrosequencing of the PCR Product

The following multistep protocol describes how the PCR products generated above are processed and how Pyrosequencing is then performed. In principle, the 5'-biotinylated primer enables the PCR products to be bound to streptavidin-coated Sepharose beads, immobilizing them during subsequent washing steps. A denaturation step is then performed to isolate the biotinylated (single) strand that was generated from the 5' biotinylated primer to serve as the template strand for the subsequent Pyrosequencing reaction. The nested Pyrosequencing primer is then annealed to this Pyrosequencing template strand before commencement of the Pyrosequencing reaction. Since Pyrosequencing is based on the sequencing-by-synthesis principle, the length of time required is dependent on the size of the sequence analyzed. For this particular assay for *MLH1* rs1799977, it takes 10 min regardless of the number of samples (approximately 1 min per nucleotide dispensation).

3.6 Immobilization of the Biotinylated PCR Product to Sepharose Beads

1. Immediately prior to using the Sepharose beads mix, ensure that the beads are thoroughly distributed in the solution by vortexing.
2. Prepare a “bead” mix comprising beads and binding buffer for each sample. Each sample requires 3 µL of the beads diluted in sufficient binding buffer (*see* Subheading 2.3 for recipe) to bring the combined volume to that of the amount of PCR reaction to be assayed. For example, for the remaining 25 µL of PCR product for each sample, this will require the addition of 22 µL binding buffer to 3 µL of beads. The beads and binding buffer can be made up as a master mix sufficient for all samples within a batch.
3. Centrifuge the PCR products to ensure no aerosols. Then aliquot 25 µL of the beads master mix into each well of the original PCR tubes/plate.
4. Mix well on a microplate shaker at 1,300 rpm (while ensuring that the liquid do not spit to the top of the tube), at room temperature for a minimum of 5 min. Constant agitation is required to keep the beads dispersed within the solution. At this stage the PCR products will now be bound to the Sepharose beads via the biotin molecule.

3.7 Strand Separation

1. Prepare the PSQ sequencing plate in advanced readiness with 40 μ L of Pyrosequencing primer diluted in annealing buffer to a concentration of 0.4 μ M into each well. Place the plate on the PSQ plate position on the PSQ 96 vacuum manifold sample prep workstation, ensuring the correct orientation with A1 on the top left corner, ensuring that it is ready for operation after the final step of the strand separation protocol.
2. Set up the PSQ 96 vacuum manifold sample prep workstation. Fill the troughs to the marked level on the side of the troughs according to Fig. 3: trough 1 contains 70 % ethanol, trough 2 denaturing buffer, and trough 3 washing buffer (see Subheading 2.3 for recipes). Trough 4 contains Milli-Q water for cleaning the vacuum tool only (not while the beads are on).
3. Clean the 96-probes vacuum prep tool by shaking it vigorously in the water trough (with the vacuum off). Then turn on the

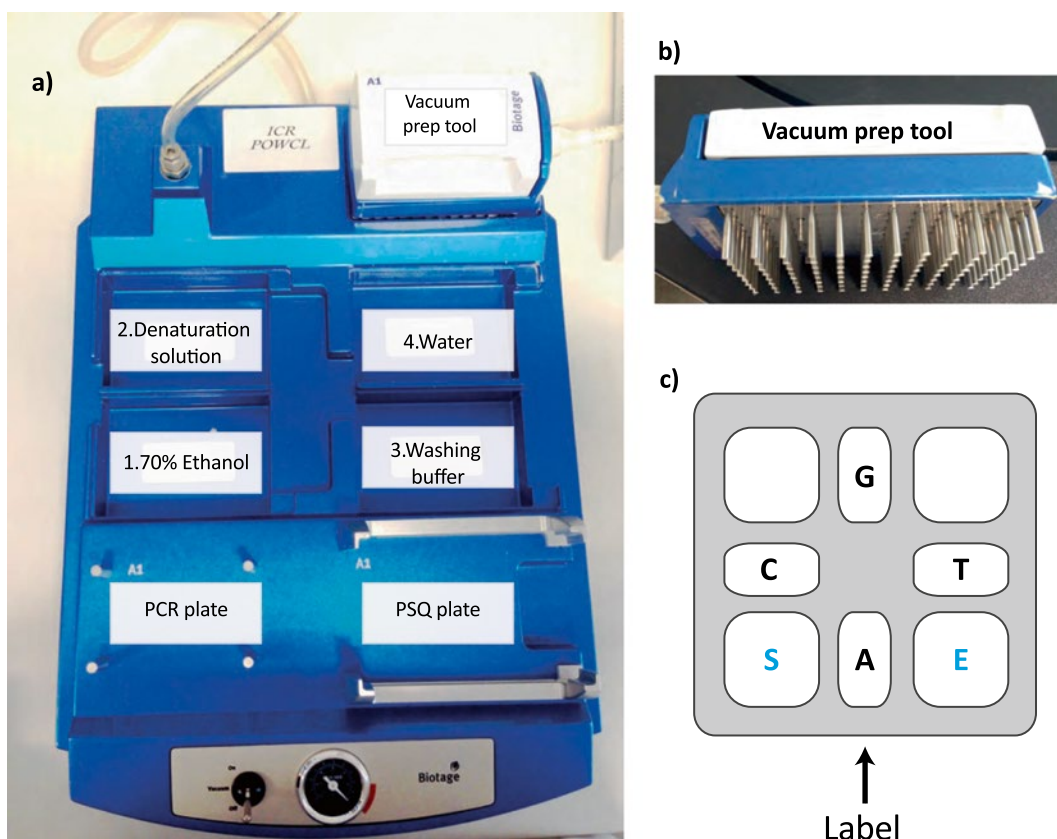


Fig. 3 Diagram of the major components of the Pyrosequencing procedure. (a) Picture of PSQ 96 vacuum manifold sample prep workstation showing the placement of the solution, PCR plate, PSQ plate and the vacuum prep tool. (b) A side-on view of the 96 probe vacuum prep tool. (c) A diagram of the PyroMark 96 cartridge showing the correct placement of nucleotides (A, C, G, T) as well as the enzyme (E) and substrate (S)

vacuum to ensure all water is aspirated from the tips of the probes (*see Note 1*).

4. Remove the PCR tubes/plate with the PCR product bound beads from the shaker and place them into the PCR plate area on the sample prep workstation (as illustrated in Fig. 3) according to the experiment plan for each sample and immediately continue to **step 5** so that the beads are still in suspension.
5. Place the vacuum tool into the mix of PCR products-beads (such that one prong of the vacuum tool is beneath the surface of the liquid in each tube/well). Very gently move the probes up and down several times to mix the beads before turning on the vacuum to ensure the beads are still in suspension. Turn on the vacuum tool. Make sure that all liquid has been aspirated from the PCR tubes/plate and all the beads have been captured on the prongs of the vacuum tool (*see Note 2*).
6. With the vacuum kept on throughout, gently remove the tool from the PCR tubes/plate to ensure that the captured beads (with double-stranded PCR products still bound) stay attached to the probe. Gently place the tool vertically into trough 1 containing 70 % ethanol for 5 s, then lift the vacuum tool to allow complete aspiration of the ethanol solution.
7. Next place the vacuum tool into trough 2 containing the denaturing buffer for 5 s. This step renders the PCR products single stranded and only the bead-bound biotinylated strand remains. Once again, lift the vacuum tool out and allow complete removal of the denaturing buffer off the prongs.
8. Finally, place it into trough 3 containing washing buffer for 5 s and lift the vacuum tool out and allow complete removal of the buffer.
9. After ensuring that the washing buffer has been completely aspirated, move the vacuum probe away from the troughs in close vicinity to the PSQ plate containing the pyrosequencing primer (which was prepared in advance in **step 1**) into which the beads will be emptied. Turn the vacuum off, then place the tool into the PSQ plate as soon as possible afterwards to minimize the possibility of bead loss (*see Note 3*).
10. After placing the probe into the sequencing primer keeping A1 on the top left corner (contained in the wells of the receiving PSQ plate), shake the tool to release the beads into the primer mix (*see Note 4*).

3.8 Annealing of the Sequencing Primer

For the annealing of the template strand onto the sequencing primer.

1. Place the PSQ plate, which now contains the biotinylated single stranded Pyrosequencing template and pyrosequencing primer, onto a heating block and heat for 2 min at 90 °C (*see Note 5*).

2. After 2 min, remove the plate from the heating block and allow it to cool slowly to room temperature. During this cooling period, the pyrosequencing primer will anneal to the single-stranded template when the appropriate annealing temperature is met.

3.9 Reconstitution of the Reagents

The PyroMark Gold Q96 Reagents including the nucleotides (dATP α S, dCTP, dGTP, and dTTP), lyophilized enzyme mixture, and the lyophilized substrate mixture are stored at 4 °C until reconstituted.

Reconstitute the lyophilized enzyme and substrate mixture with 620 μ L of Milli-Q water each, leave at room temperature for 5–10 min. Once reconstituted, store any remaining enzyme and substrate at –20 °C for future use. Enzyme and substrate stored frozen will need to be thawed to room temperature prior to use. It is advised that the frozen reagents should not go through more than three freeze–thaw cycles. Do NOT freeze the nucleotides.

3.10 Prepare the Pyrosequencing Run Using PyroMark ID Software

When using this (or any new assay) for the first time, the Pyrosequencing assay needs to be entered into the PyroMark ID program. Once entered and saved, the file containing the assay description can be quickly selected for reuse with each subsequent batch of samples, as described below (*see Note 6*).

For any new allele quantification assay the initial set up to enter a new simplex SNP entry is as follows:

1. Open the PyroMark ID program.
2. Choose “Simplex Entries.”
3. Enter the desired name for the assay in “Entry ID.”
4. Enter the “Sequence Name” and/or “Primer” if desired.
5. Enter the sequence to be analyzed with the SNP into “Sequence to analyze,” *see Fig. 2* for example of the *MLH1* rs1799977 A>G SNP.
6. Press “Dispensation order.”
7. The program will automatically generate a dispensation order.
8. Press “Histogram” to see the predicted unique histogram for each polymorphism.
9. Rename the “Position name” if desired.
10. Check for warning and errors and rectify when appropriate.
11. Save the assay.

3.11 Starting the Pyrosequencing Assay

1. Set up the Pyrosequencer for the run by using the appropriate program. Use the PyroMark ID for genotyping, allelic expression, and loss of heterozygosity studies.
2. Use the PyroMark ID program in which the assay description has been previously entered and saved.

3. Start a new run under the “SNP Run.”
4. Select the wells to be used and activate them under the “general tab.”
5. Set up the plate layout according to your sample layout.
6. Pick/enter the appropriate sample ID name or number under the “setup” tab and highlight the appropriate wells.
7. Enter the entry tab from the setup tab, select “simplex” entry and select the appropriate SNP assay for each sample.

3.12 Filling the PSQ 96 Reagent Cartridge

1. Once the correct sample number and assay has been selected, the PyroMark ID program calculates the amount of each nucleotide, enzyme, and substrate that will be required on your behalf.
2. The amount of reagents required is listed under “File.”
3. Load the nucleotides in the appropriate wells, as shown in Fig. 3, with cartridge oriented so that the label is facing towards you.
4. Load the appropriate amount of enzyme and substrate into the appropriate wells in the cartridge and avoid bubbles forming during pipetting.

3.13 Starting the Run

1. Place the PSQ plate with the notches fitting the machine to ensure the correct orientation (keeping well A1 on the top left corner at all times) and the reagent cartridge into the Pyrosequencer with the label facing you.
2. Under the general tab, select the appropriate instrument to activate the connection between the computer and the Pyrosequencer, and hit run. The Run button is only enabled when an instrument connection has been configured, the instrument parameters have been defined and assay details are defined (e.g., sample ID and simplex entry).
3. Once the run has completed, clean the PSQ 96 reagent cartridge as soon as possible. This can be achieved by filling and emptying the cartridge three times with Milli-Q water.
4. Then fill it again to the brim and press the opening of each well ten times, each time checking that the water squirts out from the pin below, repeat for all the wells used for each reagent and top the water up to the brim if necessary. A prolonged wait will increase the chance of the nucleotides drying inside the needle, which can lead to blockage in subsequent runs. Any blockage will result in defective nucleotide dispensation and hence sub-optimal runs. If there is suspicion that the cartridge is blocked, it can be soaked in Milli-Q water overnight to dissolve any deposits, and the procedure above repeated.
5. Further information on how to set up a run using the PyroMark Software and its analysis can be found in PyroMark ID instrument Reference Manual.

3.14 Raw Data Analysis

For genotyping, expression, and loss-of-heterozygosity analyses use the allelic quantification (AQ) mode of the PyroMark ID software.

1. Select the AQ mode (allelic quantification).
2. In the well overview mode, under the analyse box, select either all or selected wells for analysis. Select any of the desired information in the SNP analysis view. These include Entry ID, Sequence to analyse, Dispensation order, Sample ID, Notes, and AQ Analysis results, including Result, Quality warnings, and Analysis criteria.
3. A Pyrogram® for each well will be displayed. The SNP site will be highlighted, and the genotype will be shown in a different color, depending on the quality assessment of the SNP analysis. A blue box indicates a result that has passed quality control, yellow indicates a result that requires checking, and red indicates a failed sample.

3.15 Data Interpretation

Representative Pyrograms are shown in Fig. 4 for genotyping, loss-of-heterozygosity, and allelic expression studies at the *MLH1* c.655A>G SNP in clinical samples. Genotyping showed individual peaks with heights of 100 % for either the A allele or the G allele, in homozygotes for the opposite allele, and two peaks of similar height in heterozygotes (Fig. 4).

To detect loss-of-heterozygosity (LOH) in tumor DNA, a normal genomic DNA sample (e.g., blood or more appropriately normal tissue adjacent to the tumor) was run alongside the tumor DNA sample. The following calculation to determine the “LOH index” was performed, using the peak height data acquired for each allele of the SNP:

$$\text{LOH index} = \left(\text{Tumor}_{\text{A allele}} / \text{Tumor}_{\text{G allele}} \right) \div \left(\text{Normal}_{\text{A allele}} / \text{Normal}_{\text{G allele}} \right)$$

LOH is defined when the LOH index is <0.6 or >1.7 as previously described [10]. This means one allele in the tumor has at least a 40 % reduction in levels compared to the normal tissue. An example of LOH detected in the primary colorectal carcinoma sample from a patient with Lynch syndrome, as compared to the paired normal colorectal mucosa tissue sample from the same patient, is given in Fig. 5. This example demonstrated that, in this particular patient who had a germ-line mutation of *MLH1*, LOH served as the “second hit” in tumorigenesis [8].

To determine if there is a reduction or loss of transcription from one allele, termed an allelic expression imbalance, the relative levels of allelic expression were calculated from the peak height data for each allele in cDNA templates normalized against the level of each allele in the paired genomic DNA template, since this should eliminate any possible bias towards either allele in the assay.

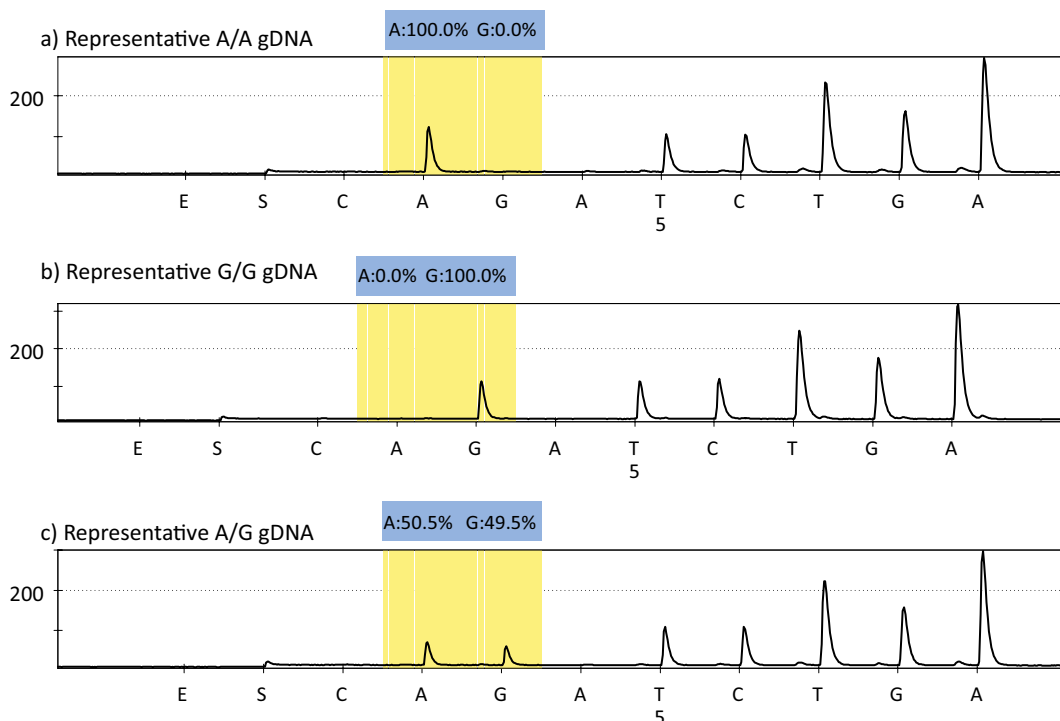


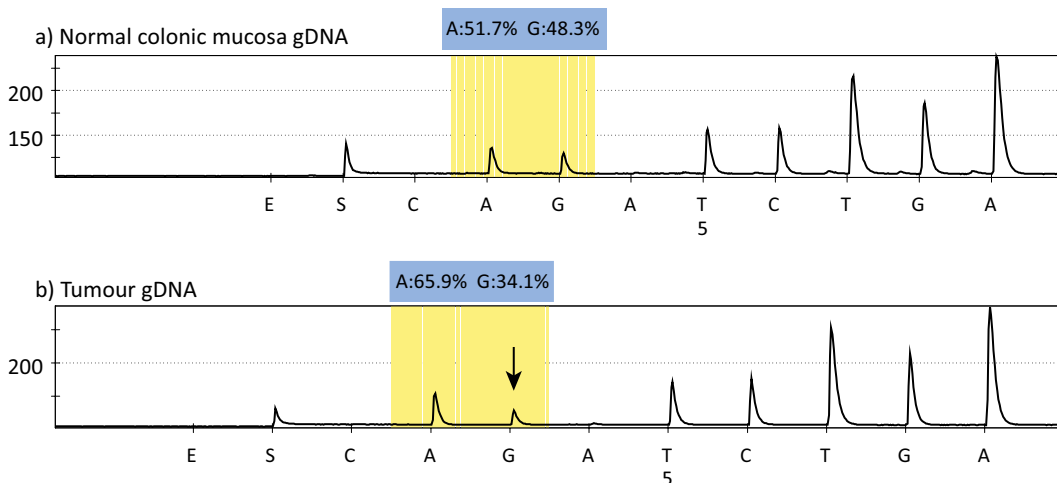
Fig. 4 Representative Pyrograms for the *MLH1* c.655A>G Pyrosequencing assays of genomic DNA. Representative output traces (Pyrograms) from genomic DNA (gDNA) showing the genotype of healthy control individuals, including (a) an A/A with a single peak of a 100 % A at the SNP site, (b) a G/G homozygote with a single peak of a 100 % G and (c) an A/G heterozygote with nearly equivalent levels of A and G allele. The SNP investigated in the assay is highlighted in yellow with the blue box over it indicating the percentage of each allele present. The height of each peak indicates the amount of each nucleotide incorporated and thus represents the relative level of each respective allele

The normalized relative levels of allelic expression were calculated as follows, and expressed as an allelic expression ratio (AER):

$$\text{AER} = \left(\text{cDNA}_{\text{A allele}} / \text{cDNA}_{\text{G allele}} \right) \div \left(\text{Genomic DNA}_{\text{A allele}} / \text{Genomic DNA}_{\text{G allele}} \right)$$

Where the AER equals ~1.0, this shows that expression levels from each allele were equivalent, giving balanced allelic expression. The AER of the *MLH1* transcripts measured at the c.655A>G SNP among four healthy heterozygous control individuals, ranged from 0.92 to 1.15 [9], with examples from two healthy controls shown in Fig. 6. Where the AER deviates significantly above or below 1.0, this indicates a loss of expression derived from one of the two alleles, as illustrated in Fig. 7 (top).

The patient illustrated in this figure has Lynch syndrome due to a constitutional *MLH1* epimutation, which is characterized by monoallelic methylation of the *MLH1* promoter throughout normal somatic tissues. This defect has caused near complete loss



$$\text{LOH index} = (\text{Tumour A allele} / \text{Tumour G allele}) \div (\text{Normal A allele} / \text{Normal G allele})$$

$$= (65.9 / 34.1) \div (51.7 / 48.3) = 1.81$$

$$\text{LOH Index (A:G)} = 1.81$$

Fig. 5 Loss of heterozygosity (LOH) detected in the tumor compared to the normal colonic mucosa from a patient with colon cancer. Pyrograms of the *MLH1* c.655A>G SNP in genomic DNA (gDNA) extracted from formalin-fixed paraffin embedded (FFPE) normal colonic mucosa shows heterozygosity as both A and G alleles were approximately equally represented (a), while the FFPE colorectal tumor specimen showed loss of heterozygosity, with a significant reduction of the G allele. The LOH index calculated from the ratio of each allele in the gDNA from tumor compared to normal tissue (1.81) is more than 1.70, indicating a loss of heterozygosity in the tumor (b)

of transcription of one allele of *MLH1* within peripheral blood cells [11]. In some cases, a partial loss of allelic expression was observed, as illustrated in Fig. 7 (bottom). This is due, in this particular patient, to the presence of a mosaic *MLH1* epimutation in which there was allele-specific methylation but not all copies of the affected allele were methylated [12].

4 Conclusion

In this chapter we have described and provided detailed protocols for allele quantification assays using Pyrosequencing that we designed to a specific SNP within the *MLH1* gene. Furthermore, we have provided some illustrative examples of the application of these assays to clinical specimens. Our interest in the *MLH1* gene relates to its role in the causation of the cancer predisposition condition Lynch syndrome. We applied these Pyrosequencing assays to aid in unravelling novel mechanisms that may affect the transcriptional function of this gene, as well as in facilitating the identification of cryptic mutations in patients with suspected Lynch

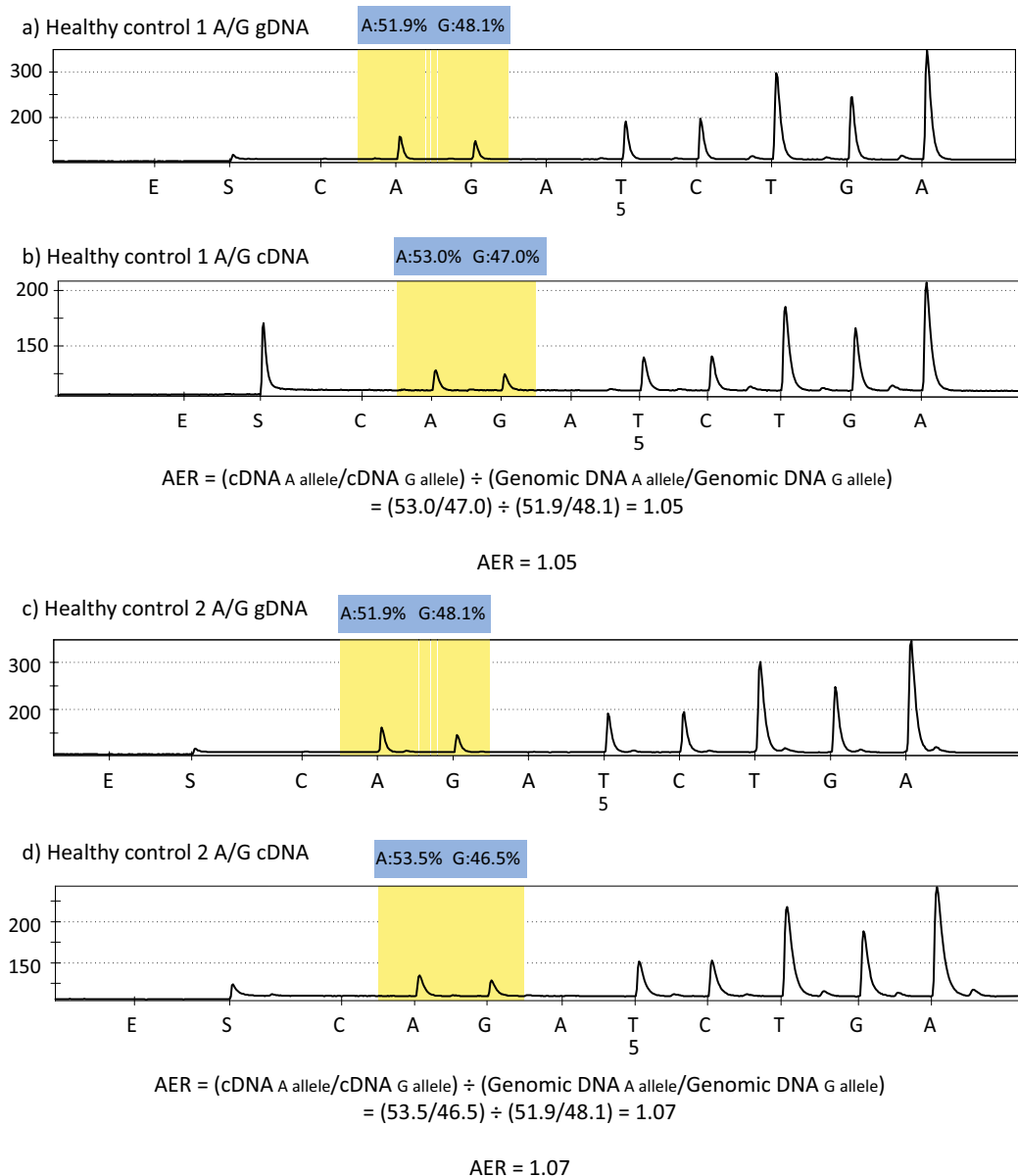


Fig. 6 Representative Pyrograms for the *MLH1* c.655A>G Pyrosequencing assays of genomic DNA and cDNA in healthy controls showing no allelic expression loss. Representative Pyrograms of the cDNAs (**b, d**) with their respective genomic DNA (**a, c**) from two healthy heterozygous individuals showing balanced biallelic transcription from both alleles. The allelic expression ratio (AER) index is close to 1

syndrome whose tumors had demonstrated MLH1 protein loss. We selected a particular benign exonic SNP as the target for allele quantification Pyrosequencing assays because it is common (frequently heterozygous in our study population) and therefore highly informative. However, we have similarly performed allele quantification Pyrosequencing assays at a private single nucleotide

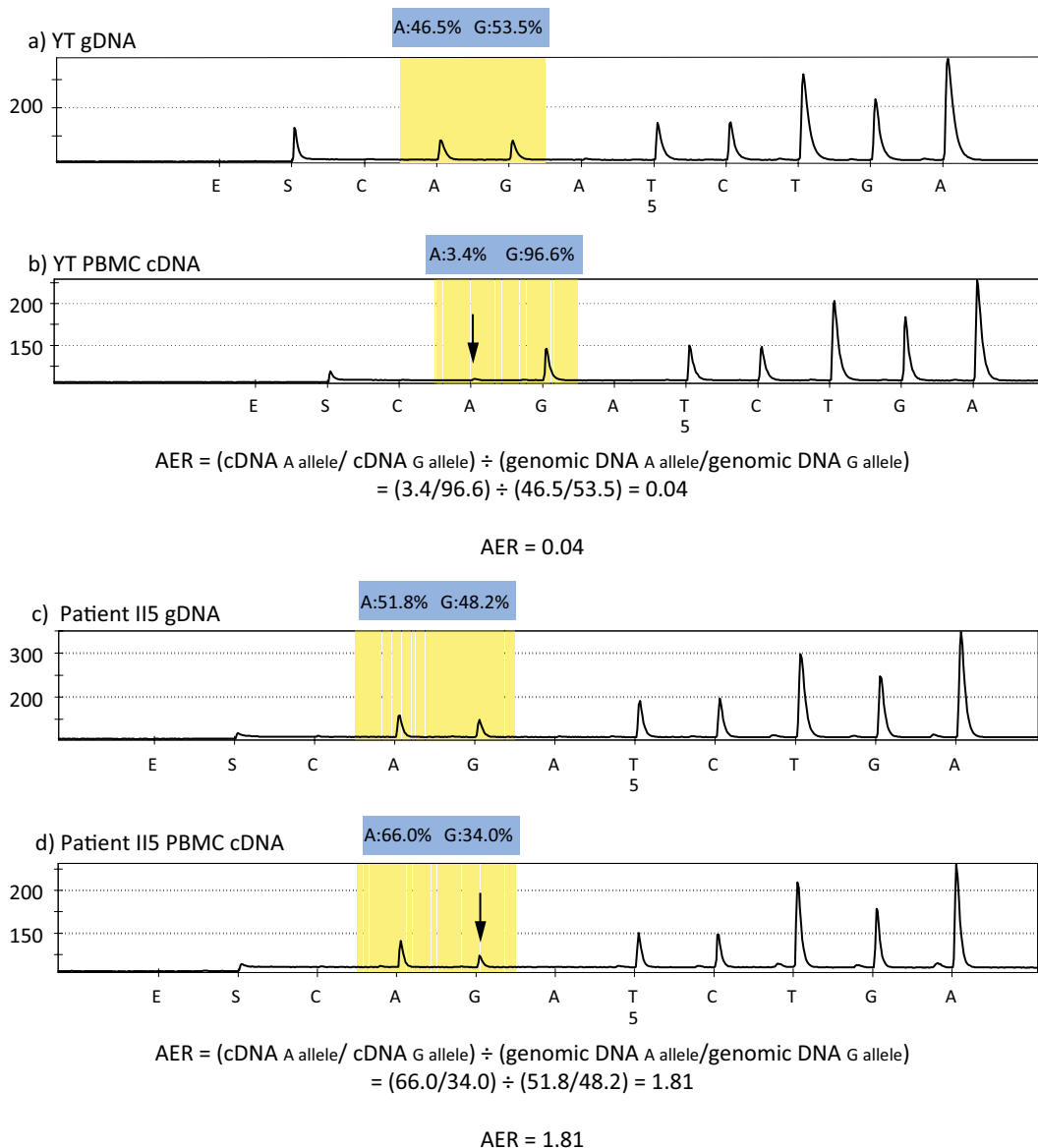


Fig. 7 Monoallelic loss of transcription of the A allele at *MLH1* c.655A>G in the PBMC by Pyrosequencing. Pyrograms across the c.655A>G SNP show the genotype and the transcription level of each allele in two patients with constitutional epimutations who are heterozygous at this SNP (a, c). These two patients show different degrees of loss of expression. Patient (YT) showing complete inactivation of the A allele in the peripheral blood mononuclear cells (PBMC) indicated by the loss of the A peak (downward arrow) (b). While patient II5 showing a partial loss of the G allele in the PBMC as shown by the significant reduction of the G peak (downward arrow) (d)

sequence variant site within the *MLH1* gene in other Lynch syndrome families who carried this variant [12, 13]. This type of assay may be custom-designed to virtually any SNP or mutation within a gene of interest to study allelic expression levels as they

occur naturally or in a pathological setting. Another advantage of Pyrosequencing is that it is a robust assay for use on clinical specimens including archival samples, because the PCR amplification products are short and Pyrosequencing does not make use of fluorescent dyes. We successfully obtained results studying loss-of-heterozygosity within tumor DNAs extracted from archival formalin-fixed paraffin embedded clinical specimens. Allele quantification Pyrosequencing assays remain our method of choice for clinically relevant investigations of allelic imbalances with pathological consequences.

5 Notes

1. Beads can accumulate on the tips of the probe, which may interfere with the capturing of new beads. Thus, it is important to ensure that any remaining beads from prior usage be dislodged prior to starting a new experiment.
2. In some instances, the liquid in the PCR tube may not be sufficiently aspirated. If this occurs, the probe should be thoroughly cleaned by vigorously agitating it in Milli-Q water. If the blockage remains, that particular probe tip should be replaced.
3. If the vacuum tool is mistakenly kept on at this stage, the primer mix will be lost through the vacuum.
4. After the beads are released into the primer solution, some may remain in the PCR tube. If the amount is significant, it is possible to add additional binding buffer (approximately 50 μ L) to the PCR tubes to resuspend the beads and restart the process after cleaning the probe properly.
5. The heating block should be preheated to 90 °C in advance to save time.
6. If you are using the PyroMark ID system, it is advisable to switch on the Pyrosequencer prior to setting up the assay. It is recommended to allow 90 min for the CCD camera on the PyroMark ID to stabilize.

References

1. Ahmadian A, Gharizadeh B, Gustafsson AC et al (2000) Single-nucleotide polymorphism analysis by pyrosequencing. *Anal Biochem* 280:103–110
2. Ronaghi M (2001) Pyrosequencing sheds light on DNA sequencing. *Genome Res* 11:3–11
3. Wang L, Baudhuin LM, Boardman LA et al (2004) MYH mutations in patients with attenuated and classic polyposis and with young-onset colorectal cancer without polyps. *Gastroenterology* 127:9–16
4. Packham D, Ward RL, Ap Lin V et al (2009) Implementation of novel pyrosequencing assays to screen for common mutations of BRAF and KRAS in a cohort of sporadic colorectal cancers. *Diagn Mol Pathol* 18:62–71

5. Reik W, Walter J (2001) Genomic imprinting: parental influence on the genome. *Nat Rev Genet* 2:21–32
6. Ferguson-Smith AC (2011) Genomic imprinting: the emergence of an epigenetic paradigm. *Nat Rev Genet* 12:565–575
7. Lee JT, Bartolomei MS (2013) X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell* 152:1308–1323
8. Knudson AG Jr (1971) Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* 68:820–823
9. Kwok CT, Ward RL, Hawkin NJ et al (2010) Detection of allelic imbalance in MLH1 expression by pyrosequencing serves as a tool for the identification of germline defects in Lynch syndrome. *Fam Cancer* 9:345–356
10. Canzian F, Salovaara R, Hemminki A et al (1996) Semiautomated assessment of loss of heterozygosity and replication error in tumors. *Cancer Res* 56:3331–3337
11. Goel A, Nguyen TP, Leung HC et al (2011) De novo constitutional MLH1 epimutations confer early-onset colorectal cancer in two new sporadic Lynch syndrome cases, with derivation of the epimutation on the paternal allele in one. *Int J Cancer* 128:869–878
12. Hitchins MP, Rapkins RW, Kwok CT et al (2011) Dominantly inherited constitutional epigenetic silencing of MLH1 in a cancer-affected family is linked to a single nucleotide variant within the 5'UTR. *Cancer Cell* 20:200–213
13. Kwok CT, Vogelaar IP, van Zelst-Stams WA et al (2013) The MLH1 c.-27C>A and c.85G>T variants are linked to dominantly inherited MLH1 epimutation and are borne on a European ancestral haplotype. *Eur J Hum Genet* 22:617–624

Free ebooks ==> www.Ebook777.com

Part III

Analysis of DNA Methylation Patterns

Free ebooks ==> www.Ebook777.com

Chapter 13

Quantitative DNA Methylation Analysis by Pyrosequencing®

Jessica Roessler and Ulrich Lehmann

Abstract

DNA methylation is an epigenetic mark playing an important role in development and disease. Aberrant DNA methylation was identified as an alternative mechanism for gene inactivation complementing deletions and mutations in cancer initiation and progression. However, to accurately compare differences in DNA methylation among various tissue types, adequate quantitative approaches are required. Pyrosequencing®, as a sequencing-by-synthesis method, allows such quantification with single CpG resolution and the ability for threshold determination. This book chapter provides a detailed protocol for DNA methylation analysis by Pyrosequencing, including information on assay design and practical procedure. Additionally, emphasis is placed on the discussion of strengths and weaknesses of the methodology.

Key words DNA methylation, Pyrosequencing®, Quantification, Assay design

1 Introduction

The term DNA methylation denotes the addition of a methyl group on the fifth position of a cytosine. Although methylation marks on non-CpG sites were also found [1, 2], the vast majority of methylated cytosines are followed by guanine (CpG). During cellular development but also during disease initiation and/or progression changes in DNA methylation have been found to play a major role. Cancer cells for instance, exhibit a global loss of DNA methylation accompanied by an aberrant hypermethylation of CpG-rich regions, so called CpG islands, within the promoter region of especially tumor suppressor genes [3]. Therefore, the DNA methylation status of promoter regions of specific genes, such as *MLH1* or *MGMT*, is already an established marker in routine cancer diagnostics [4, 5].

Most of the methods developed so far for the investigation of DNA methylation are based on the treatment with sodium bisulfite [6, 7]. This implies that each cytosine, which is not protected by a methyl group gets deaminated into uracil (thymine after PCR), whereas methylated cytosines remain unaltered. By this means the

epigenetic information gets translated into the underlying primary DNA sequence information and can be regarded as a C/T polymorphism.

Fast, easy-to-use, and cost-effective methods to screen for DNA methylation patterns are COBRA [8] and methylation-specific PCR (MSP) [9]. Yet, COBRA is limited by the number of restriction sites within the region of interest. The major drawback of MSP is the high rate of false positives and severe limitations in the analysis of CpG-poor regions. The discrimination of methylated against unmethylated DNA is based on the inclusion of as many CpG sites as possible in the primer binding sites. However, the discrimination of insufficiently converted genomic DNA requires an appropriate number of non-CpG sites within the primer region. These requirements are difficult to achieve. In addition to these limitations, COBRA and MSP also lack the ability to quantify the results, delivering only information about presence or absence of DNA methylation marks.

Pyrosequencing[®] is an easy-to-use, fast, and reproducible method to analyze not only CpG-rich but also CpG-poor regions [10]. It is a quantitative sequencing-by-synthesis approach, making it possible to define a threshold discriminating hypomethylation or hypermethylation and/or differences in DNA methylation among various tissue types (like tumor versus normal tissue). Our laboratory for example defines the threshold for a significant difference among tumor and normal tissue by the mean of the DNA methylation level of normal tissue plus two times the standard deviation [11]. However, one has to keep in mind that the limit of detection for Pyrosequencing is usually around 3–5 %, depending on the respective assay ([12, 13], *see also* Chapter 4). The major disadvantage of the method is the limited read length of a single Pyrosequencing reaction. Depending on the region of interest only a small number of CpG sites can be analyzed (on average 5–10 CpG sites).

Within the last years an increasing number of genome-wide DNA methylation methods have been developed. One of these new methods is the 450 K Infinium[®] BeadArray from Illumina[®] Inc. [14]. It is a bead based microarray approach delivering beta-values (range 0–1) being comparable to the percent of DNA methylation determined by Pyrosequencing (0–100 %). The cross-validation of 450 K BeadArray and Pyrosequencing results exhibited a high congruence among these two methods for DNA methylation analysis [15]. Nevertheless, although the 450 K BeadArray is a feasible approach to investigate DNA methylation changes between tissue subtypes, it is still too expensive for most laboratories to screen large numbers of samples. Therefore, Pyrosequencing could be the method of choice to verify candidate regions identified by genome-wide screening methods within a significantly larger number of patient samples. Besides, an independent validation of candidate

regions identified by genome-wide methods will always be required and Pyrosequencing is becoming the standard reference method for DNA methylation approaches [15, 16].

2 Materials

2.1 DNA Isolation

1. DNeasy Blood and Tissue Kit (Qiagen®, Hilden, Germany).
2. Xylol.
3. Ethanol.
4. Proteinase K with buffer: 50 mM Tris-HCl, pH 8.1, 1 mM EDTA pH 8.0, 0.5 % Tween20.
5. Phenol-chloroform-isoamylalcohol.
6. Chloroform.
7. Isopropanol.
8. RNase A.
9. Sodium acetate: 3 M, pH 7.0 with 100 µg/mL dextran.
10. TE-buffer: 10 mM Tris-HCl, pH 8.1, 1 mM EDTA pH 8.0.
11. Overhead rotator.
12. Thermomixer (up to 1,400 rpm).
13. Tabletop centrifuge (up to 16,200 × g).

2.2 Bisulfite Treatment

1. EZ DNA Methylation™ Kit (ZymoResearch, Irvine, CA, USA).
2. Ethanol.
3. Tabletop centrifuge (up to 16,200 × g).
4. Heating block with cover to ensure light protection.

2.3 PCR

1. Taq-Polymerase including buffer and MgCl₂.
2. dNTPs.
3. Forward and reverse primer (*see* Subheading 3.1 Assay Design).
4. PCR tubes or plates.
5. Thermocycler.

2.4 Pyrosequencing

1. Pyrosequencing system (Qiagen, Hilden, Germany) (*see* Note 1).
2. Pyro Q-CpG™ software (Qiagen).
3. Reagent cartridges (Qiagen).
4. Capillary cartridges (Qiagen).
5. Dispensing unit/Cartridge holder (Qiagen).
6. Pyrosequencing reagent kit (Qiagen).
7. Primers for sequencing reaction.

8. Biotinylated PCR product.
9. Streptavidin Sepharose® HP (GE Healthcare, Chalfont St Giles, GB).
10. PCR plates (96-well plate using the PyroMark®Q96 ID system).
11. Plate mixer (100 up to 2,000 rpm).
12. Pyrosequencing plates (PyroMark Q96 Plate Low).
13. Vacuum workstation with the appropriate filter tips and troughs (Qiagen).
14. PyroMark Binding buffer (Qiagen).
15. Denaturation solution: 0.2 N NaOH.
16. Washing buffer concentrate (Qiagen).
17. Annealing buffer (Qiagen).
18. High purity water.
19. Ethanol.
20. Heating block (heating up to 80 °C) with PyroMark Q96 Sample Prep Thermoplate Low (Qiagen).
21. PSQ Assay design software.
22. Sequence Conversion Tool.

3 Methods

3.1 Assay Design

There are four different possibilities to design a DNA methylation Pyrosequencing® assay for a specific region. Since the methylation status of one region is the same within the Watson and Crick strand, both of these can be used as template. Additionally, the sequencing can be performed in both directions, meaning forward or reverse. For the design of Pyrosequencing assays Biotage® provided the PSQ Assay design software with an updated version now being available from Qiagen. Although there are also various other free software tools available (*see Note 2*), the PSQ Assay design software offers the possibility to individually change primer position and length, generating an instant response in respect to melting temperature and secondary structure formation of the single primer or within the assay itself. However, the program has been designed for the SNP analysis application of Pyrosequencing. Since bisulfite-treated DNA templates are of reduced complexity, due to the conversion of unmethylated cytosines into thymidines, potential errors or complications for Pyrosequencing assays are more frequently notified by the software.

Although the ideal primer for the amplification reaction does not include any CpG site (methylation site), this prerequisite is hard to obtain by analyzing CpG sites located within CpG-rich

regions, such as CpG islands. Next to that, bisulfite-treated DNA is of reduced complexity (easily transformed by the Sequence Conversion Tool [17], *see Note 2*), making the primer design even more complicated. It is therefore recommended to place the primer in such a way, that the CpG site(s) is (are) within the 5' region of the primer, with at least five nucleotides at the 3' region harboring no potential methylation site. Methylation sites within the primer should be designed with degenerated bases, Y (C/T) for the forward direction and R (G/A) for the reverse direction. However, SNPs or other polymorphisms existing within the genomic region of interest could potentially bias the sequencing reaction (for a worst case scenario *see* [18]) and should be avoided. To increase the specificity of the oligonucleotides for fully converted DNA, PCR primers should ideally harbor a minimum of four and the sequencing primer at least one cytosine of a non-CpG site.

Otherwise the same rules as for normal PCR primer design have to be considered. The GC content should not exceed 60 %, the annealing temperature should be less than 65 °C, long sequences of homopolymers as well as repetitive sequences should be avoided, thus defining a unique primer binding site. The size of the amplification product should be of maximal 350 bp (for FFPE tissue samples the amplicon length should not exceed 150 bp, since FFPE tissue derived DNA is usually highly fragmented) to avoid secondary structure formation within the single stranded template. Such secondary structures could be the reason for an increased background signal or even the inhibition of the sequencing reaction.

Once the amplification primers are positively tested the assay has to be set up. The Pyro Q-CpG software requires the insertion of the sequence to analyze and generates the dispensation order for the assay on its own. The dispensation order is a predetermination of nucleotide incorporation, whereas next to nucleotides expected to be incorporated also nucleotides which are expected to give no signal are included (called blank dispensation). Each generated signal which is not associated with a CpG site (methylation detection site) is called a reference peak, used as internal control concerning quality issues and to calculate the expected single peak height. The histogram function represents the theoretical peak height. Internal controls like these and conversion controls are the major advantage of DNA methylation analysis by Pyrosequencing. As conversion control, thymines (forward direction) or adenines (reverse direction) are suggested by the software to control the bisulfite-mediated reaction. These genomic sites are required to be converted during bisulfite treatment, meaning thymines have been cytosines (forward direction) and adenines have been guanines (reverse direction) within the unconverted genomic sequence.

The duration of each run can be calculated by the number of nucleotides to be dispensed and the expected cycle time of 65 s.

3.2 DNA Isolation

1. Resuspend fresh-frozen tissue sections in about 750 μL Proteinase K-buffer and 25 μL Proteinase K.
2. Incubate at 55 $^{\circ}\text{C}$ and 500 rpm overnight.
3. Add 750 μL phenol–chloroform–isoamylalcohol and mix by inverting for 30 min.
4. Centrifuge at room temperature for 10 min at $16,200\times g$.
5. Recover the aqueous phase and repeat **steps 3** and **4** until no interphase is visible anymore.
6. Add 50 μg RNase A and incubate for 30 min at 37 $^{\circ}\text{C}$ and 800 rpm.
7. Repeat **steps 3** and **4**, to remove the RNase A.
8. Add 750 μL chloroform to the aqueous phase, mix by inverting 30 min.
9. Subsequently centrifuge for 3 min, $16,200\times g$ at room temperature.
10. Precipitate the DNA adding 0.1 volumes of 3 M sodium acetate (pH 7.0)/100 $\mu\text{g}/\text{mL}$ dextran and 2.5 volumes 100 % ethanol and inversion of the solution.
11. Wash the precipitate with 500 μL of 70 % ethanol and dissolve it within TE buffer.

3.3 Bisulfite Treatment

Depending on DNA quality and amplification results of the PCR, 0.4–1 μg genomic DNA is bisulfite-treated using the EZ DNA MethylationTM Kit (*see Note 3*).

1. Mix 5 μL M-Dilution Buffer and the respective amount of sample DNA, adjusted with distilled water to a total volume of 50 μL and incubate at 42 $^{\circ}\text{C}$ for 15 min.
2. Add 100 μL of CT Conversion Reagent (dissolved using 750 μL distilled water and 210 μL M-Dilution buffer) to each sample and mix. Subsequently incubate at 50 $^{\circ}\text{C}$ for 12–16 h in the dark.
3. Incubate on ice for 10 min.
4. Place the Zymo-SpinTM IC Column into a provided Collection Tube.
5. Mix each sample with 400 μL of M-Binding Buffer and transfer the solution to the Zymo-SpinTM IC Column.
6. Centrifuge for 30 s at full speed ($>10,000\times g$). Discard the flow-through.
7. Add 100 μL M-Wash Buffer and centrifuge at full speed for 30 s.
8. Add 200 μL M-Desulfonation Buffer and incubate at room temperature for 15 min. Subsequently, centrifuge for 30 s at full speed.

9. Add 200 μ L M-Wash Buffer, centrifuge at full speed for 30 s and repeat this step one more time.
10. Place the column into a 1.5 mL microcentrifuge tube. Elute the converted DNA by adding about 40 μ L of M-Elution Buffer directly to the column matrix and centrifugation at full speed for 30 s.

3.4 Pyro-PCR

1. PCR reactions amplifying bisulfite-treated DNA for subsequent Pyrosequencing analysis are performed using 0.4 μ M non-tailed, 0.4 μ M universal biotinylated (*see Note 4*), and 0.04 μ M tailed primer. The preparation of primer master mixes is recommended ensuring the presence of all three primers within the PCR reaction. Examples for each of such primer, corresponding to the Pyrograms[®] in Fig. 2, are shown here:

Tailed forward primer: ggg ACA CCgCTg ATC gTT TAT YgggTT TTT TAT Ygg ATT ATT Ag.

Reverse primer: CAA CRC TAC CRC TAC TAC TAA TTT.

Universal biotinylated primer: Biotin-ggg ACA CCgCTg ATC gTT TA.

2. The optimal $MgCl_2$ concentration as well as the annealing temperature has to be defined individually for each primer set.
3. Depending on the predetermined PCR conditions 1.5 mM or 2.5 mM $MgCl_2$ (*see step 2*) were mixed with 0.2 mM dNTP, 0.02 units/ μ L Platinum[®] Taq polymerase (*see Note 5*), and the primer master mix.
4. The PCR should last for at least 45 cycles, ensuring a sufficient amount of template to be generated and to avoid undesired background due to competing biotinylated primer.
5. Subsequent gel electrophoresis should give a strong and distinct band of the PCR product with minimal primer contaminations and unspecific byproducts.

3.5 Pyrosequencing

1. Switch on the heating block (80 °C) with the thermoplate placed onto it.
2. Start the PyroMark Q96 ID system; a 90 min warm-up to stabilize the output of the CCD camera is recommended.
3. Start the Pyro Q-CpG software. Generate the plate setup by importing the assay file (either by dragging it in or right-click) and denoting sample IDs. Under the tab “Tools-> Volume Information” the system calculates the required amount of enzyme, substrate, and nucleotides, later on necessary for loading the cartridges.
4. Fill the troughs of the Vacuum workstation with
 - (a) 70 % v/v ethanol.
 - (b) 0.2 N NaOH.

- (c) Washing buffer.
 - (d) Distilled water.
5. Mix the following reagents and agitate them at 1,400 rpm in a 96-well PCR plate at room temperature for 5 min (*see Note 6*):
 - (a) 47 μL binding buffer.
 - (b) 3 μL Streptavidin Sepharose® beads (mix gently but well before use).
 - (c) Up to 20 μL PCR product.
 - (d) Fill up to a total volume of 80 μL with high purity water.
 6. For each sequencing reaction add 12 μL of 0.83 μM sequencing primer in annealing buffer into each respective well of the Pyrosequencing plate. The samples positioned within the PCR plate have to be placed at exactly the same position within the Pyrosequencing plate, thus the sequencing primer will be able to anneal to the appropriate PCR product. To prevent any misplacement, it is advisable to perform the plate setup before starting the procedure (*see step 3*).
 7. Apply vacuum to the Vacuum workstation and wash the filter tips for 10 s in the trough containing distilled water (*see Note 7*).
 8. Lower the filter probes into the PCR plate and aspirate the liquid containing immobilized PCR product (takes about 15 s). Cautiously take out the aspiration device of the plate.

Be careful not to loose the Streptavidin Sepharose beads!
 9. The Streptavidin Sepharose bead coupled DNA fragments are now placed on top of the filter of the filter tips, visible as small white hills (Fig. 1).
 10. To wash the beads place the aspiration device for 5 s into 70 % ethanol (trough 1). Subsequently, the double stranded DNA will be denatured by placing the filter tips into trough 2 (denaturation solution) for 10 s. The beads will be washed again using washing buffer (trough 3) for additional 10 s. Be careful not to touch the trough with the filter tips!
 11. Remove all residual liquids in the aspiration device by raising and tilting it 90°, such that all liquids within the filter tips can drain off.
 12. Place the filter tips directly above the Pyrosequencing plate, turn off the vacuum, and wait until the pressure is back to normal. Lower the aspiration device into the Pyrosequencing plate, thus immerse the filter tips into the annealing buffer containing the sequencing primer and release the Streptavidin Sepharose beads into the solution by gently shaking the aspiration device.

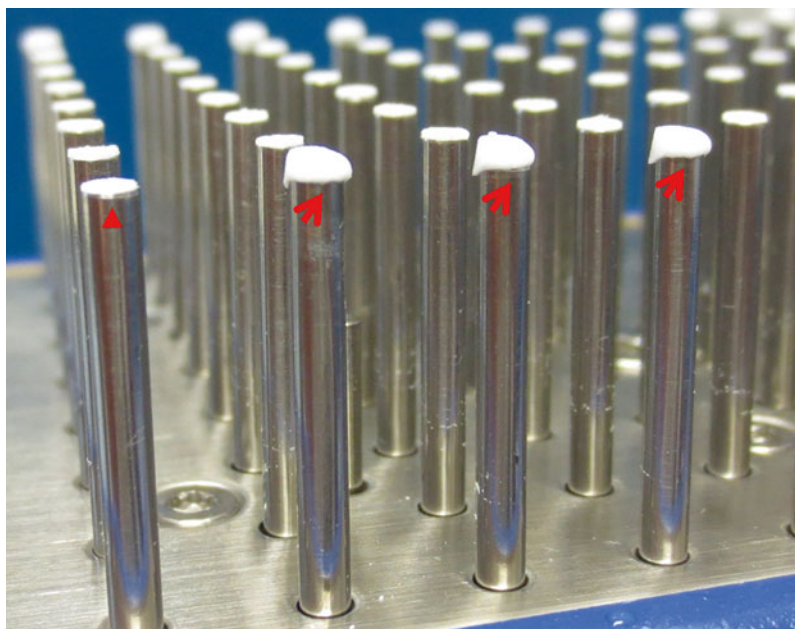


Fig. 1 Filter tips of the Pyrosequencing aspiration device. Streptavidin Sepharose beads are visible as small white hills on top of the filter membrane of the filter tips, marked by *arrows* (*arrowhead*: filter tip without Streptavidin Sepharose bead on top)

13. To anneal the sequencing primer, place the Pyrosequencing plate on the preheated heating block (80 °C) and cover the sequencing plate with the Thermoplate. Incubate for 2 min and subsequently allow the samples to cool down to room temperature for at least 5 min.
14. Centrifuge the nucleotides for 5 min, $13,000\times g$, at 4 °C to settle any kind of particles, thus avoiding their transfer into the cartridges. These particles could occlude the cartridge, potentially preventing correct dispensation of the affected nucleotide.
15. As stated in the “Volume information,” apply the required amount of enzyme, substrate, and nucleotide into each respective cartridge. Take care of air bubbles since these could lead to dispensation errors as well. Place the cartridges in the respective position of the cartridge holder.
16. Position the cartridge holder and the loaded Pyrosequencing plate into the PyroMark Q96 ID, close the lid, and start the run via the Pyro Q-CpG software.
17. Right after the run has finished carefully clean the cartridges with distilled water to avoid occlusion of the cartridge tip (*see Note 8*).

3.6 Pyrogram Evaluation

The outcome of the sequencing reaction is displayed in a so-called Pyrogram, where relative light units (RLU) are plotted against a time scale, indicated by nucleotide dispensation. The DNA methylation status is displayed at the top of each Pyrogram, directly above each individual CpG site analyzed. This percentage value represents the average methylation level of all sequenced PCR products for the respective CpG site investigated within the assay.

Peaks of a well running assay are distinct and sharp, reaching the expected heights (as indicated in the histogram function, Fig. 2).

A decrease in peak heights, especially for long assays with increasing running time, is characteristic due to declining enzyme activity and dilution effects caused by an increase in liquid volume per well. Since the number of incorporated nucleotides is proportional to the emission of light, the same amount of consecutive

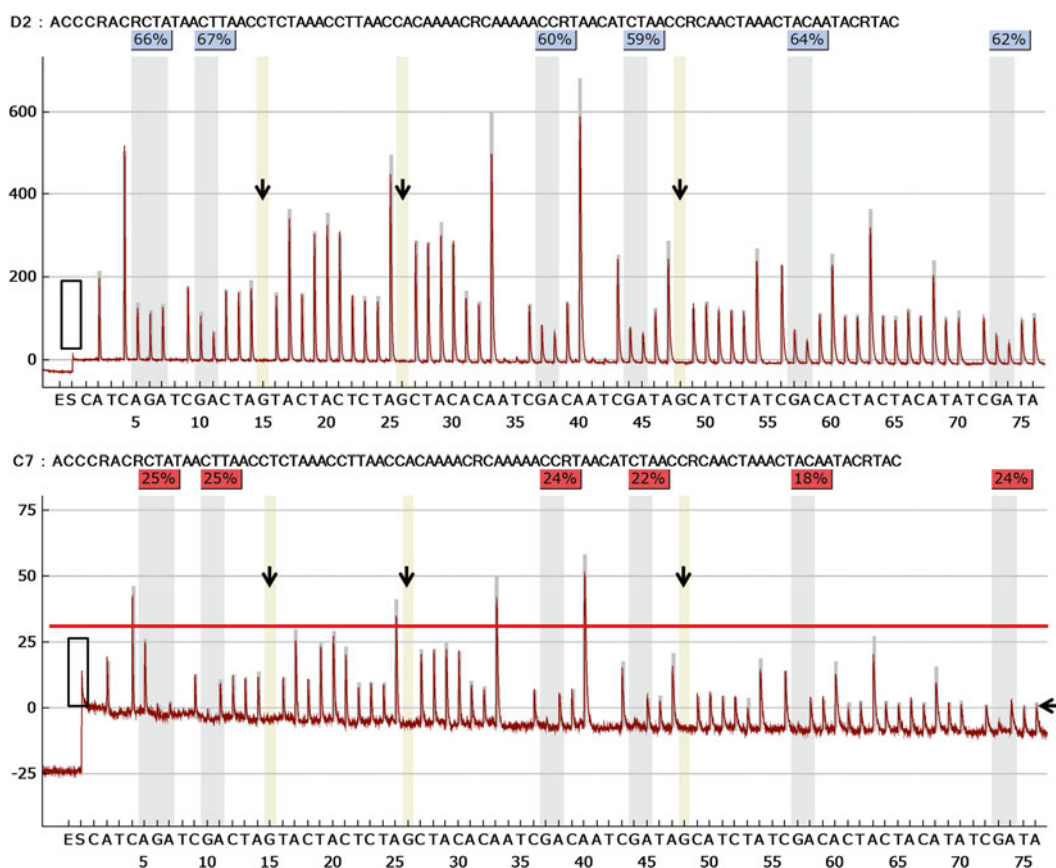


Fig. 2 Pyrosequencing runs of different quality. Figure (a) shows a high-quality Pyrosequencing run, with distinct and sharp peaks within the expected intensity height. The bisulfite conversion control is lacking an intensity signal, as expected. Figure (b) represents a Pyrosequencing run of bad quality exhibiting intensity signals of single nucleotides barely exceeding 25 RLU (red line). Next to that, high background signals and a decline of the baseline can be observed (horizontal arrow head). CpG sites are marked by a gray bar, conversion controls by a yellow bar and vertical arrow heads. Histogram function is represented by a gray rectangular shadow behind each peak. Black boxes at the beginning of the Pyrograms mark the substrate peak

peaks should give similar peak heights for “C,” “G,” and “T,” with peaks for “A” being slightly higher. Nevertheless, with increasing numbers of homopolymers this expected peak heights will not be reached properly.

As already mentioned above, the Pyro Q-CpG software delivers a quality control system for each sequencing run. Conversion controls of completely converted genomic DNA at a specific locus should not show any intensity signal. The threshold for these conversion control signals to pass the quality criteria can be set manually for each assay template. Default settings of the software for this threshold to pass quality control (blue assay) is a percentage of 4.5 and a percentage of 7.0 for conversion controls that have to be rechecked manually (yellow colored). Ready-to-use assays available from Qiagen usually have these quality criteria set to a percentage of 6.5 for passing and 9.5 for rechecking this control.

Blank dispensations should not exhibit any intensity signal being significantly above background signals (*see Note 9*).

The signal intensity of the expected peaks should be strong enough to be distinguished from background signals, therefore peak heights of at least 30–50 RLU are preferable. The judgment of this threshold should be considered in the context of the other quality criteria described herein. In case of inappropriate peak heights the system will usually give a warning message concerning a low signal-to-noise ratio.

From time to time, reference peaks seem to be displaced, meaning they are detected before the expected position within the histogram. This type of error occurs mainly due to dispensation mistakes and can be removed by excluding the respective reference peak. Additionally, repeat the run using a new cartridge.

A large substrate dispensation peak (the first potentially visible peak) would indicate a contamination of the reaction mixture with pyrophosphate (Box in Fig. 2).

4 Notes

1. Qiagen is the sole supplier for Pyrosequencing analysis. Thus, the sequencer, the Vacuum workstation with the respective equipment, cartridges and cartridge holder, and most of the reagents can be purchased only from Qiagen or its local suppliers.
2. Software suitable for primer design of DNA methylation analysis:

MethPrimer (based on Primer3 algorithm) [19]: <http://www.urogene.org/methprimer/>.

MethylPrimer Express®: <http://www.appliedbiosystems.com/absite/us/en/home/support/software-community/free-ab-software.html>.

Bisearch [20]: <http://bisearch.enzim.hu/>.

Biotage PSQ Assay design: <http://psq-assay-design.software.informer.com/>.

Sequence Conversion Tool [17]: <http://biq-analyzer.bioinf.mpi-inf.mpg.de/tools/BiConverter>.

3. Our laboratory performs bisulfite treatment by means of the EZ DNA methylation kit from Zymo Research. Numerous other conversion kits are commercially available as well. No judgment on the reliability and reproducibility of any other kits can be given herein, however, positive experience was gained by using the EZ DNA Methylation Kit. Nevertheless, the use of the Zymo Research kit is suggested for genome-wide DNA methylation screening methods, such as the 450 K BeadArray of Illumina Inc., indicative of a reliable means for bisulfite conversion.
4. Assays designed for Pyrosequencing analysis requires one of the primers to be biotinylated. However, biotinylated primers are less stable than unlabeled primers, with frequent freezing-thawing cycles being detrimental. Biotinylated primers are required to have a high purity (HPLC purified) to avoid cross-reactions of unbound biotin with the Streptavidin Sepharose beads. An option to reduce the cost per sequencing assay is the use of a universal biotinylated primer [21]. The sequence of this primer, carrying the biotin-mark, is the same as the “tail” sequence added to one of the amplification primers (for detail *see* Chapter 8). Although this is a convenient approach for most assays, one should keep in mind that it can also interfere with the amplification or even sequencing reaction.
5. The Platinum Taq DNA polymerase from Invitrogen is the DNA polymerase of choice we are using in the laboratory. However, DNA polymerases specifically suitable for GC-rich regions have been developed within the last years. Qiagen even offers a ready-to-use Pyrosequencing PCR kit including the Q-solution, which is supposed to optimize the amplification of difficult PCR templates such as GC-rich regions.
6. The Streptavidin Sepharose beads settle quite fast, therefore the PCR plate containing the immobilized PCR product should be agitated right until capturing. In case of elapsing more than 1 min, agitate the plate again before proceeding.
7. The PyroMark Q96 ID User Manual suggests changing the filter tips of the aspiration device after approx. 100 runs. Therefore, one can test the reliability of each tip by aspirating a 96-well plate completely filled with high-purity water. Wells exhibiting large amount of residual water are indicative of filter tips which should be replaced.
8. Cartridges can be reused several times (20–30 times). Therefore, they have to be cleaned carefully right after each

run by discarding residuals within the cartridge and rinsing it three times with high-purity water. To avoid occlusion of the cartridge tip, completely fill the cartridge with high-purity water and press on top of the cartridge with a finger, a jet of water should come out of the tip. Before further use, let the cartridge dry on a lint-free cloth. Additionally, a dispensation test can be performed right before each run. Therefore, a Pyrosequencing plate is sealed with foil and placed at its respective position within the instrument. After the loaded cartridge holder is inserted as well, the dispensation test can be initiated by the respective icon in the software. If six droplets are visible on the foil at the expected positions, the cartridges are working fine. In case droplets are missing, replace the respective cartridge.

9. In exceptional cases, reference peaks with no expected signal show distinct and sharp peaks as well. One possible explanation for these extra peaks could be a single nucleotide polymorphism within the respective tissue sample. However, it could be as well the result of lacking congruence with the reference genome sequence (hg18/hg19). Additionally, an out-of-phase sequencing reaction is also possible. That means, some templates fail to incorporate a respective nucleotide compared to the rest of templates (minus shift) or some templates incorporate more than one nucleotide at a dispensation cycle, thus being ahead of the residual templates (plus shift). This could be the result of suboptimal sequencing primers or their binding conditions.

References

1. Lister R, Pelizzola M, Dowen RH et al (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462:315–322
2. Pelizzola M, Ecker JR (2011) The DNA methylome. *FEBS Lett* 585:1994–2000
3. Kondo Y, Issa JP (2010) DNA methylation profiling in cancer. *Expert Rev Mol Med* 12:e23
4. Lehmann U (2010) DNA methylation. From basic research to routine diagnostics. *Pathologie* 31(Suppl 2):274–279
5. Ogino S, Kawasaki T, Brahmandam M et al (2006) Precision and performance characteristics of bisulfite conversion and real-time PCR (MethyLight) for quantitative DNA methylation analysis. *J Mol Diagn* 8:209–217
6. Laird PW (2010) Principles and challenges of genome-wide DNA methylation analysis. *Nat Rev Genet* 11:191–203
7. Zilberman D, Henikoff S (2007) Genome-wide analysis of DNA methylation patterns. *Development* 134:3959–3965
8. Eads CA, Laird PW (2002) Combined bisulfite restriction analysis (COBRA). *Methods Mol Biol* 200:71–86
9. Herman JG, Graff JR, Myöhänen S (1996) Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci U S A* 93:9821–9826
10. England R, Pettersson M (2005) Pyro Q-CpG™: quantitative analysis of methylation in multiple CpG sites by pyrosequencing. *Nat Methods* 2:i–ii
11. Lehmann U, Berg-Ribbe I, Wingen LU et al (2005) Distinct methylation patterns of benign and malignant liver tumors revealed by quantitative methylation profiling. *Clin Can Res* 11:3654–3660

12. Harrington CT, Lin EI, Olson MT et al (2013) Fundamentals of pyrosequencing. *Arch Pathol Lab Med* 137:1296–1303
13. Tost J, Gut IG (2007) DNA methylation analysis by pyrosequencing. *Nat Protocols* 2:2265–2275
14. Bibikova M, Barnes B, Tsan C et al (2011) High density DNA methylation array with single CpG site resolution. *Genomics* 98: 288–295
15. Roessler J, Ammerpohl O, Gutwein J et al (2012) Quantitative cross-validation and content analysis of the 450 k DNA methylation array from Illumina, Inc. *BMC Res Notes* 5:210
16. Potapova A, Albat C, Hasemeier B et al (2011) Systematic cross-validation of 454 sequencing and pyrosequencing for the exact quantification of DNA methylation patterns with single CpG resolution. *BMC Biotechnol* 11:6
17. Bock C, Reither S, Mikeska T et al (2005) BiQ analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. *Bioinformatics* 21:4067–4068
18. Tost J, Jammes H, Dupont JM et al (2007) Non-random, individual-specific methylation profiles are present at the sixth CTCF binding site in the human H19/IGF2 imprinting control region. *Nucl Acids Res* 35:701
19. Li L-C, Dahiya R (2002) MethPrimer: designing primers for methylation PCRs. *Bioinformatics* 18:1427–1431
20. Tusnady GE, Simon I, Varadi A et al (2005) BiSearch: primer-design and search tool for PCR on bisulfite-treated genomes. *Nucl Acids Res* 33:e9
21. Colella S, Shen L, Baggerly KA et al (2003) Sensitive and quantitative universal pyrosequencing methylation analysis of CpG sites. *Biotechniques* 35:146–150

Chapter 14

Quantitative Methylation Analysis of the *PCDHB* Gene Cluster

Barbara Banelli and Massimo Romani

Abstract

Long Range Epigenetic Silencing (LRES) is a repressed chromatin state of large chromosomal regions caused by DNA hypermethylation and histone modifications and is commonly observed in cancer. At 5q31 a LRES region of 800 kb includes three multi-gene clusters (*PCDHA@*, *PCDHB@*, and *PCDHG@*, respectively). Multiple experimental evidences have led to consider the *PCDHB* cluster as a DNA methylation marker of aggressiveness in neuroblastoma, second most common solid tumor in childhood. Because of its potential involvement not only in neuroblastoma but also in other malignancies, an easy and fast assay to screen the DNA methylation content of the *PCDHB* cluster might be useful for the precise stratification of the patients into risk groups and hence for choosing the most appropriate therapeutic protocol. Accordingly, we have developed a simple and cost-effective Pyrosequencing® assay to evaluate the methylation level of 17 genes in the protocadherin B cluster (*PCDHB@*). The rationale behind this Pyrosequencing assay can in principle be applied to analyze the DNA methylation level of any gene cluster with high homologies for screening purposes.

Key words Pyrosequencing®, DNA methylation, *PCDHB@* cluster, Cancer, Multiplex

1 Introduction

1.1 Background

Pyrosequencing® is considered the gold standard technology for targeted quantitative DNA methylation analysis. It consists of the sequence-by-synthesis reaction of short sequences of genomic DNA, which have been modified with sodium bisulfite and then amplified with sequence specific primers. This chemical modification converts only unmethylated Cytosines into Uracils, leaving the methylated Cytosines unchanged thus translating the presence of DNA methylation into a nucleotide difference (C indicates the presence of a methyl group in the genomic DNA while T means absence of methylation).

The region of interest, which usually includes many CpG sites, is amplified with a pair of primers, one of which biotinylated

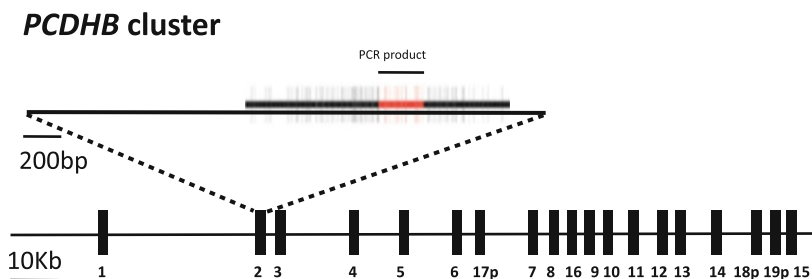


Fig. 1 Genomic structure of the *PCDHB* gene cluster. The *black vertical bars* represent the 16 genes (1–16) and the three pseudogenes (17p–19p) of *PCDHB* cluster. The schematic expansion of *PCDHB* 2 is representative of the structure of each *PCDHB* member. The CpG island located within the single exon of each gene is indicated by a *thick line*. The region analyzed by Pyrosequencing is indicated in *red*

at the 5' end. The biotinylated single strand of the amplified fragments is then analyzed by Pyrosequencing using a sequencing primer.

The peaks of light resulting from the Pyrosequencing reaction at the “methylation variable positions” created through the bisulfite conversion are translated into percentage of methylation by the instrument's software.

Long Range Epigenetic Silencing (LRES) is a repressed chromatin state of large chromosomal regions caused by DNA hypermethylation and histone modifications and is commonly observed in cancer. At 5q31 a LRES region of 800 kb includes three multi-gene clusters (*PCDHA@*, *PCDHB@*, and *PCDHG@*, respectively). Multiple experimental evidences have led to consider the *PCDHB* cluster as a DNA methylation marker of aggressiveness in neuroblastoma, second most common solid tumor in childhood [1–3]. Because of its potential involvement not only in neuroblastoma but also in other malignancies, an easy and fast assay to screen the DNA methylation content of the *PCDHB* cluster might be useful for the precise stratification of the patients into risk groups and hence for choosing the most appropriate therapeutic protocol.

1.2 Principle of the Assay

The assay analyzes the mean DNA methylation level of the *PCDHB* cluster. The *PCDHB* cluster consists of 16 genes and three pseudogenes, all highly conserved and each one hosting a CpG Island (Fig. 1).

The assay described in the following protocol permits the unbiased simultaneous amplification and sequencing of 15 genes and 2 pseudogenes of the *PCDHB* cluster for quantitative methylation analysis, taking into account all sequence variations.

Gene 1 and pseudogene 19P were excluded from the analysis, as their homology with the remaining *PCDHB* members is not sufficient to include them in the same assay as noted previously [2].

2 Materials

2.1 DNA Extraction and Quality Assessment

1. DNeasy Blood & Tissue Kit (Qiagen).
2. Picodrop Microliter UV/Vis Spectrophotometer.

2.2 Bisulfite Modification

1. Thermal cycler.
2. EpiTect Bisulfite Kit (Qiagen).
3. Minicentrifuge.

2.3 PCR Reaction and Gel Control of the PCR Products

1. Pyrosequencing Assay Design Software (Qiagen).
2. UV-cabinet for PCR operations.
3. PCR and sequencing primers.
4. IMMOLASE Hot Start DNA Polymerase (Bioline).
5. Thermal cycler.
6. UltraPure™ Agarose-1000.

2.4 Pyrosequencing Reaction

1. PSQ 96MA instrument (Qiagen).
2. Vacuum Prep Tool™ (Qiagen).
3. PyroMark Gold Q96 Reagents.
4. Streptavidin Sepharose™ High Performance (GE Healthcare).
5. Binding buffer (Qiagen).
6. Annealing buffer (Qiagen).
7. PSQ 96 Plate Low™ (Qiagen).
8. PyroQ-CpG™ software (Qiagen).

3 Methods

The *PCDHB* methylation Pyrosequencing assay consists of a multiplex PCR reaction to amplify 17 *PCDHB* genes and a subsequent Pyrosequencing reaction with three consecutive sequencing primers to obtain the mean methylation levels of 20 CpG sites. The basic principles of the reaction are:

1. Fragments of the 17 *PCDHB* genes are amplified with the same primer pair in a single PCR reaction.
2. The sequencing primers anneal to the amplified fragments of all considered *PCDHB* genes.
3. The Pyrosequencing reaction proceeds on all genes in real time and along all the established sequences without interruption.

3.1 Alignment of the *PCDHB* Gene Sequences and Primers Design

Genes and pseudogenes of *PCDHB* cluster were aligned and after in silico bisulfite modification, all nucleotide variations were noted (Fig. 2) (*see Note 1*).

The PCR primer pair was designed with the Pyrosequencing Assay Design Software (Qiagen) considering the regions of minor variability among all the *PCDHB* genes to amplify the *PCDHB* genes in the same PCR reaction. A single degenerate base (V) was introduced in the reverse PCR primer to take a single nucleotide variation (T/C/G) among the 17 sequences into account (Table 1 and Fig. 3) (*see Note 2*).

3.2 Sequencing Primer Design

The PCR reaction provides a number of amplification products that retain in their sequences the nucleotide differences of the various *PCDHB* genes. We designed three sequencing primers to sequence a large part of the PCR amplification product. Every nucleotide difference of the *PCDHB* gene was considered during the design of the sequencing primers: in two of the three sequencing primers some degenerate bases were introduced to enable the simultaneous Pyrosequencing of all genes (Table 1).

3.3 Assessment of DNA Quality and Integrity

1. Prior to the bisulfite modification, verify the DNA integrity by electrophoresis on a 1 % agarose gel.
2. Measure the DNA quality spectrophotometrically using the 260 nm/280 nm ratio and calculate the DNA concentration from the 260 nm measure.

3.4 Bisulfite Modification

1. Modify 500 ng of DNA with sodium bisulfite to convert the unmethylated Cytosines (but not the methylated Cytosines) into Uracils using the EpiTect Bisulfite kit and following the manufacturer's instructions.
2. Perform the final elution step with 40 μ L of the elution buffer.

3.5 Generation of the *PCDHB* Amplicons

1. Set up the PCR reactions in triplicate so that every PCR product can be analyzed by Pyrosequencing with one of the three sequencing primers within the same 96-well plate (*see Note 3*).
2. Prepare a PCR master Mix. For a 50 μ L reaction volume use 0.3 μ M of forward and biotinylated-reverse PCR primer, 0.2 mM of each dNTP, 1.5 mM of $MgCl_2$, 1 $\times$ Taq reaction buffer, and 1.25 units IMMOLASE Hot Start DNA Polymerase.
3. Generate the amplification products using the following PCR conditions: 95 °C for 10 min, followed by 45 cycles with denaturation at 95 °C for 30 s, annealing at 59 °C for 30 s, and elongation at 72 °C for 30 s, and a final elongation cycle at 72 °C for 10 min.
4. Verify 10 μ L of the 253 bp PCR products on a 3 % agarose gel.

```

1          20          40          60          80
PCDHB_2 : GGGTTATTTGGTGATTAAAGTGGTGGCGGTGGACGGCGATTTCGGGTTAGAACGTTTGGTTGTCGTATTAGTTGTTTAAAGGTTACG
PCDHB_3 : GGGTTATTTGGTGATTAAAGTGGTGGCGGTGGACGGCGATTTCGGGTTAGAACGTTTGGTTGTCGTATTAGTTGTTTAAAGGTTACG
PCDHB_4 : GGGTTATTTGGTGATTAAAGTGGTGGCGGTGGACGGCGATTTCGGGTTAGAACGTTTGGTTGTCGTATTAGTTGTTTAAAGGTTACG
PCDHB_5 : GGGTTATTTGGTGATTAAAGTGGTGGCGGTGGACGGCGATTTCGGGTTAGAACGTTTGGTTGTCGTATTAGTTGTTTAAAGGTTACG
PCDHB_6 : GGGTTATTTGGTGATTAAAGTGGTGGCGGTGGACGGCGATTTCGGGTTAGAACGTTTGGTTGTCGTATTAGTTGTTTAAAGGTTACG
PCDHB_7 : GGGTTATTTGGTGATTAAAGTGGTGGCGGTGGACGGCGATTTCGGGTTAGAACGTTTGGTTGTCGTATTAGTTGTTTAAAGGTTACG
PCDHB_8 : GGGTTATTTGGTGATTAAAGTGGTGGCGGTGGACGGCGATTTCGGGTTAGAACGTTTGGTTGTCGTATTAGTTGTTTAAAGGTTACG
PCDHB_9 : GGGTTATTTGGTGATTAAAGTGGTGGCGGTGGACGGCGATTTCGGGTTAGAACGTTTGGTTGTCGTATTAGTTGTTTAAAGGTTACG
PCDHB10 : GGGTTATTTGGTGATTAAAGTGGTGGCGGTGGACGGCGATTTCGGGTTAGAACGTTTGGTTGTCGTATTAGTTGTTTAAAGGTTACG
PCDHB11 : GGGTTATTTGGTGATTAAAGTGGTGGCGGTGGACGGCGATTTCGGGTTAGAACGTTTGGTTGTCGTATTAGTTGTTTAAAGGTTACG
PCDHB12 : GGGTTATTTGGTGATTAAAGTGGTGGCGGTGGACGGCGATTTCGGGTTAGAACGTTTGGTTGTCGTATTAGTTGTTTAAAGGTTACG
PCDHB13 : GGGTTATTTGGTGATTAAAGTGGTGGCGGTGGACGGCGATTTCGGGTTAGAACGTTTGGTTGTCGTATTAGTTGTTTAAAGGTTACG
PCDHB14 : GGGTTATTTGGTGATTAAAGTGGTGGCGGTGGACGGCGATTTCGGGTTAGAACGTTTGGTTGTCGTATTAGTTGTTTAAAGGTTACG
PCDHB15 : GGGTTATTTGGTGATTAAAGTGGTGGCGGTGGACGGCGATTTCGGGTTAGAACGTTTGGTTGTCGTATTAGTTGTTTAAAGGTTACG
PCDHB16 : GGGTTATTTGGTGATTAAAGTGGTGGCGGTGGACGGCGATTTCGGGTTAGAACGTTTGGTTGTCGTATTAGTTGTTTAAAGGTTACG
PCDHB17p : GGGTTATTTGGTGATTAAAGTGGTGGCGGTGGACGGCGATTTCGGGTTAGAACGTTTGGTTGTCGTATTAGTTGTTTAAAGGTTACG
PCDHB18p : GGGTTATTTGGTGATTAAAGTGGTGGCGGTGGACGGCGATTTCGGGTTAGAACGTTTGGTTGTCGTATTAGTTGTTTAAAGGTTACG

86          100          120          140          160
PCDHB_2 : GAGTTCGGGTTGTTTCGGCGTGTGGGCGTATAATGGCGAGGTGCGTATCGTTAGGTTGTTGAGGAGCGCGACGTTGTTAAGTATA
PCDHB_3 : GAGTTCGGGTTGTTTCGGCGTGTGGGCGTATAATGGCGAGGTGCGTATCGTTAGGTTGTTGAGCGAGCGCGACGCGGTTAAGTATA
PCDHB_4 : GAGTTTCGGTTGTTTCGGCGTGTGGGCGTATAATGGCGAGGTGCGTATCGTTAGGTTGTTGAGCGAGCGCGACGTTAAGTATA
PCDHB_5 : GAGTTCGGGTTGTTTCGGCGTGTGGGCGTATAATGGCGAGGTGCGTATCGTTAGGTTGTTGAGCGAGCGCGACGCGGTTAAGTATA
PCDHB_6 : GAGTTCGGGTTGTTTCGGCGTGTGGGCGTATAATGGCGAGGTGCGTATCGTTAGGTTGTTGAGCGAGCGCGACGCGGTTAAGTATA
PCDHB_7 : GAGTTCGGGTTGTTTCGGCGTGTGGGCGTATAATGGCGAGGTGCGTATCGTTAGGTTGTTGAGCGAGCGCGACGCGGTTAAGTATA
PCDHB_8 : GAGTTCGGGTTGTTTCGGCGTGTGGGCGTATAATGGCGAGGTGCGTATCGTTAGGTTGTTGAGCGAGCGCGACGCGGTTAAGTATA
PCDHB_9 : GAGTTCGGGTTGTTTCGGCGTGTGGGCGTATAATGGCGAGGTGCGTATCGTTAGGTTGTTGAGCGAGCGCGACGCGGTTAAGTATA
PCDHB10 : GAGTTCGGGTTGTTTCGGCGTGTGGGCGTATAATGGCGAGGTGCGTATCGTTAGGTTGTTGAGCGAGCGCGACGCGGTTAAGTATA
PCDHB11 : GAGTTCGGGTTGTTTCGGCGTGTGGGCGTATAATGGCGAGGTGCGTATCGTTAGGTTGTTGAGCGAGCGCGACGCGGTTAAGTATA
PCDHB12 : GAGTTCGGGTTGTTTCGGCGTGTGGGCGTATAATGGCGAGGTGCGTATCGTTAGGTTGTTGAGCGAGCGCGACGCGGTTAAGTATA
PCDHB13 : GAGTTCGGGTTGTTTCGGCGTGTGGGCGTATAATGGCGAGGTGCGTATCGTTAGGTTGTTGAGCGAGCGCGACGCGGTTAAGTATA
PCDHB14 : GAGTTCGGGTTGTTTCGGCGTGTGGGCGTATAATGGCGAGGTGCGTATCGTTAGGTTGTTGAGCGAGCGCGACGCGGTTAAGTATA
PCDHB15 : GAGTTCGGGTTGTTTCGGCGTGTGGGCGTATAATGGCGAGGTGCGTATCGTTAGGTTGTTGAGCGAGCGCGACGCGGTTAAGTATA
PCDHB16 : GAGTTCGGGTTGTTTCGGCGTGTGGGCGTATAATGGCGAGGTGCGTATCGTTAGGTTGTTGAGCGAGCGCGACGCGGTTAAGTATA
PCDHB17p : GAGTTCGGGTTGTTTCGGCGTGTGGGCGTATAATGGCGAGGTGCGTATCGTTAGGTTGTTGAGCGAGCGCGACGCGGTTAAGTATA
PCDHB18p : GAGTTCGGGTTGTTTCGGCGTGTGGGCGTATAATGGCGAGGTGCGTATCGTTAGGTTGTTGAGCGAGCGCGACGCGGTTAAGTATA

171          180          200          220          240
PCDHB_2 : GGTTCGGTGGTGTGGTTAAGGATAATGGCGAGTTTTTCGGGTTTCGGTTATCGTTACGTTGTACGTGTTTTTGGTGAGCGGTTTTTT
PCDHB_3 : GGTTCGGTGGTGTGGTTAAGGATAATGGCGAGTTTTTCGGGTTTCGGTTATCGTTACGTTGTAGTGTGTTTTTGGTGAGCGGTTTTTT
PCDHB_4 : GGTTCGGTGGTGTGGTTAAGGATAATGGCGAGTTTTTCGGGTTTCGGTTATCGTTACGTTGTACGTGTTTTTGGTGAGCGGTTTTTT
PCDHB_5 : GGTTCGGTGGTGTGGTTAAGGATAATGGCGAGTTTTTCGGGTTTCGGTTATCGTTACGTTGTACGTGTTTTTGGTGAGCGGTTTTTT
PCDHB_6 : GGTTCGGTGGTGTGGTTAAGGATAATGGCGAGTTTTTCGGGTTTCGGTTATCGTTACGTTGTACGTGTTTTTGGTGAGCGGTTTTTT
PCDHB_7 : GGTTCGGTGGTGTGGTTAAGGATAATGGCGAGTTTTTCGGGTTTCGGTTATCGTTACGTTGTACGTGTTTTTGGTGAGCGGTTTTTT
PCDHB_8 : GGTTCGGTGGTGTGGTTAAGGATAATGGCGAGTTTTTCGGGTTTCGGTTATCGTTACGTTGTAGTGTGTTTTTGGTGAGCGGTTTTTT
PCDHB_9 : GGTTCGGTGGTGTGGTTAAGGATAATGGCGAGTTTTTCGGGTTTCGGTTATCGTTACGTTGTACGTGTTTTTGGTGAGCGGTTTTTT
PCDHB10 : GGTTCGGTGGTGTGGTTAAGGATAATGGCGAGTTTTTCGGGTTTCGGTTATCGTTACGTTGTAGTGTGTTTTTGGTGAGCGGTTTTTT
PCDHB11 : GGTTCGGTGGTGTGGTTAAGGATAATGGCGAGTTTTTCGGGTTTCGGTTATCGTTACGTTGTAGTGTGTTTTTGGTGAGCGGTTTTTT
PCDHB12 : GGTTCGGTGGTGTGGTTAAGGATAATGGCGAGTTTTTCGGGTTTCGGTTATCGTTACGTTGTACGTGTTTTTGGTGAGCGGTTTTTT
PCDHB13 : GGTTCGGTGGTGTGGTTAAGGATAATGGCGAGTTTTTCGGGTTTCGGTTATCGTTACGTTGTACGTGTTTTTGGTGAGCGGTTTTTT
PCDHB14 : GGTTCGGTGGTGTGGTTAAGGATAATGGCGAGTTTTTCGGGTTTCGGTTATCGTTACGTTGTACGTGTTTTTGGTGAGCGGTTTTTT
PCDHB15 : GGTTCGGTGGTGTGGTTAAGGATAATGGCGAGTTTTTCGGGTTTCGGTTATCGTTACGTTGTAGTGTGTTTTTGGTGAGCGGTTTTTT
PCDHB16 : GGTTCGGTGGTGTGGTTAAGGATAATGGCGAGTTTTTCGGGTTTCGGTTATCGTTACGTTGTACGTGTTTTTGGTGAGCGGTTTTTT
PCDHB17p : GGTTCGGTGGTGTGGTTAAGGATAATGGCGAGTTTTTCGGGTTTCGGTTATCGTTACGTTGTACGTGTTTTTGGTGAGCGGTTTTTT
PCDHB18p : GGTTCGGTGGTGTGGTTAAGGATAATGGCGAGTTTTTCGGGTTTCGGTTATCGTTACGTTGTACGTGTTTTTGGTGAGCGGTTTTTT

```

Fig. 2 Alignment of the 17 *PCDHB* sequences after bisulfite modification. The sequence shown is the expected sequence in the case of complete methylation and the one considered for the design of the Pyrosequencing assay. Mismatches are indicated in *light green*. Modified from ref. [2]

3.6 Pyrosequencing Assay

The Pyrosequencing assay was performed with a PSQ 96MA instrument using the PyroMark Gold Q96 Reagents according to the manufacturer's instructions.

1. Add 40 μ L of the PCR products to a mix consisting of 3 μ L Streptavidin Sepharose HP™ and 37 μ L Binding buffer
2. Mix at 1,400 rpm for 15 min at room temperature.

Table 1
PCR and sequencing primers used for the pyrosequencing assay

(a) PCR primers
Forward: GGGTTATTTGGTGATTAAGGTG
Reverse: bi3255
(b) Sequencing primers
S1 primer: TTTGGTGATTAAGGTGG
Sequence to analyze
TGGYGGTGGAYGGYGATTYGGGTTAGAAYGTTTGGTTGTYGTATTAG
Dispensation order
GTAGTCGTGTATCAGTCGTATCGTAGTATCGTGTAGTCGTA
S2 primer: YRGYRTGTGGGYGTATAAT
Sequence to analyze
GGYGAGGTGYGTATYGTTAGGTTGTTGAGYGAGYGYGAYG
Dispensation order
TGTCGAGTAGTCGTGATCGTAGTGTGATGTCGATGTCAGATCGTATC
S3 primer: TGGTGT*TKGTTAAGGATAA
Sequence to analyze
TGGYGAGTTTTYGYGTTYGGTTATYGT*TTAYGTTGTAYGTGT
Dispensation order
GTAGTCGATGTTTCAGTCTGTCGTGATCGTGATCGTGTGATCGTG

3. Use the Vacuum Prep Tool™ to prepare single-stranded PCR products (for a detailed description of this procedure *see* for example Chapter 7).
4. Release the Sepharose beads with the single-stranded templates attached into a plate PSQ 96 Plate Low™ plate containing a mix of 45 µL Annealing Buffer and 0.4 mM of the respective sequencing primer.
5. Perform the Pyrosequencing reactions in a PSQ 96MA Pyrosequencer with the PyroMark® Gold Q96 Reagents according to the manufacturer's instructions.
6. Perform the DNA methylation analysis with the PyroQ-CpG™ software, specifically dedicated to the methylation analysis of multiple CpG sites within a single target. The software requires an internal file or "Assay" consisting of the precise bisulfite-modified sequence (named "sequence to analyze"). The software then automatically generates the best nucleotides dispensation order according to the "sequence to analyze". (For a more detailed description of these steps *see* for example Chapter 13.)

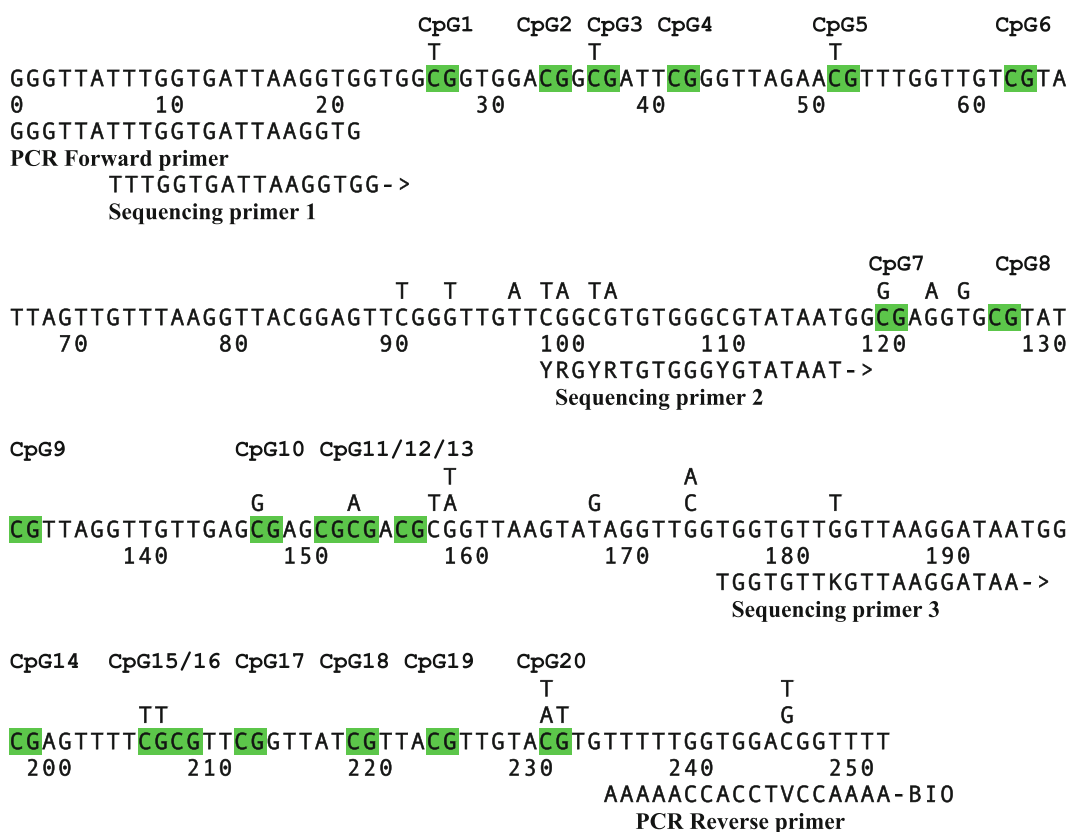


Fig. 3 Annotated sequence of the amplified *PCDHB* consensus genomic fragment after bisulfite modification considered for the design of the pyrosequencing assay. The 20 CpG doublets are highlighted in *light green* and their numbering is indicated above the sequence. The base differences between the 17 genes are reported above the sequence. The Pyrosequencing primers for amplification and sequencing are reported below the sequence. Modified from ref. [2]

In the case of the *PCDHB* cluster three assays were created: For every assay a “sequence to analyze” consisting of a nucleotide backbone common to all *PCDHB* genes, with the CpG dinucleotides indicated as “YG” was provided.

The bases that differ between the *PCDHB* genes, but which were identical to the previous or to the following base in the dispensation, were sequenced within the originally devised dispensation order without additional modifications. The only exception was the mismatch at base 154 of the PCR product, where the default dispensation order established by the Pyro Q-CpG software was modified by providing a different base (A) immediately before the next methylation variable position (Y). This additional nucleotide corresponds to nucleotide 40 of the S2 dispensation order. All nucleotide differences have to be considered as a single base difference in one of the *PCDHB* genes results in the interruption and/or the out-of-frame proceeding of the Pyrosequencing

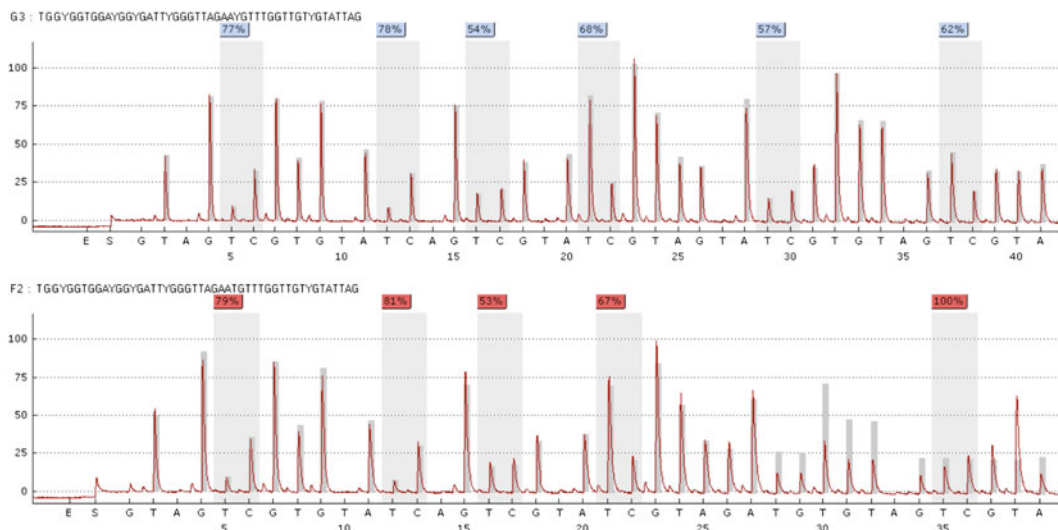


Fig. 4 Pyrograms resulted from sequencing with primer S1 of the same sample. The upper Pyrogram is correctly sequenced; the lower Pyrogram displays an error in the dispensation order, which produces a partial interruption of the sequence from dispensation 28 to 34 and successive out of frame reading

run for that gene (Fig. 4), and generate erroneous methylation values for the following YGs. Table 1 lists the sequence to analyze and the relative dispensation order for each assay.

3.7 Data Analysis

The Pyro Q-CpG software has been designed to obtain the quantitative methylation levels of the CpG sites under investigation by sequencing the amplified fragments derived from an unequivocally determined genome region.

In the case of the *PCDHB* cluster, the multiple alignments revealed that some base changes occur also within the CpG sites. The software does not consider other bases except C (methylated cytosine) and T (unmethylated cytosine prior to bisulfite conversion), but some *PCDHB* genes present a G or an A instead of the C of the CpG doublet (CpG 7, 10, 12, and 20). In addition, some *PCDHB* genes have instead of the CpG site a TpG site in the genomic sequence, which is of course not a site of potential methylation and therefore not modified by sodium bisulfite (CpG 1, 3, 5, 15, and 20). The presence of this type of mismatches hampers the direct read-out of the results by the analysis software, as this, not distinguishing between native Thymines and Thymines derived from the bisulfite conversion of unmethylated Cytosines, would alter the estimated percentage of methylation at these doublets.

To overcome these interpretation difficulties, the correct methylation level can be derived from the raw Pyrosequencing data utilizing the Peaks Heights output (*see Note 4*). The Peaks Height output records the height of the peaks of light in an Excel® file for each nucleotide dispensation.

1. Choose the reference peaks for every CpG sites to be corrected according to the following criteria:
 - (a) The reference peak should be a single base conserved in all 17 sequences.
 - (b) It could be any base, except A, as the peak height is “adjusted” by the instrument according to an internal algorithm.
 - (c) Assuming the above conditions met, the nearest possible base to the interrogated Cytosine should be chosen.
2. Attribute a value of 100 % to the reference peak and calculate the corrected percentage of methylation according to the following proportion:

Height of peak of C: % methylation = height of the reference peak: 100 %.

The CpG sites that contain sequence mismatches, the nucleotides used as reference peaks and their relative numbers in the dispensation order, are indicated in Table 2. In the Excel file resulting from the Peaks Heights output these coordinates identify the height of the peaks to be used in the mathematical formula indicated above.

A representative *PCDHB* DNA methylation assay is depicted in Fig. 5. The three Pyrograms® were obtained with sequencing primers S1, S2, and S3, respectively. Below the CpGs to be corrected, highlighted with an arrow, a table shows the height of the raw peaks and the percentage of methylation calculated as described.

4 Notes

1. The Pyrosequencing primer design applied to bisulfite-based methylation studies was performed with Pyrosequencing Assay Design Software. The software creates the PCR primer pair and the sequencing primer on an in silico bisulfite-converted template sequence obtained through the following steps:
 - (a) The “CG” dinucleotides are substituted with “YG” where Y, according to the IUPAC code, means that these bases become either C (methylated) or T (unmethylated) after bisulfite conversion.
 - (b) The remaining “C’s” are modified to “T” as these nucleotides outside the CpG dinucleotide are commonly not methylated in the genomic DNA. In the *PCDHB* cluster methylation analysis, the in silico modification was performed in the first step for all the genes separately; subsequently the resulting sequences were aligned.

Table 2
Reference peaks to correct the measured of methylation level [%] for some CpG sites

Sequencing primer	CpG doublets to be corrected	Reference peaks (underscored) to correct % of methylation of some CpG sites (highlighted green)	Number of dispensation of the C peak according to the dispensation order	Number of dispensation of the reference peak according to the dispensation order
S1	CpG1	TGGCGGT <u>G</u> GA	6	8
	CpG3	GCGAT <u>T</u>	17	18
	CpG5	TAGAA <u>C</u> GTTT	30	31
S2	CpG7	GGCGAGGTG <u>C</u> GT	4	13
	CpG10	GAGCGAG	31	26
	CpG12	GAGCG <u>C</u> GA	42	43
S3	CpG15	TTCGCG <u>T</u> T	13	19
	CpG20	GTACGT <u>G</u> TT	41	44

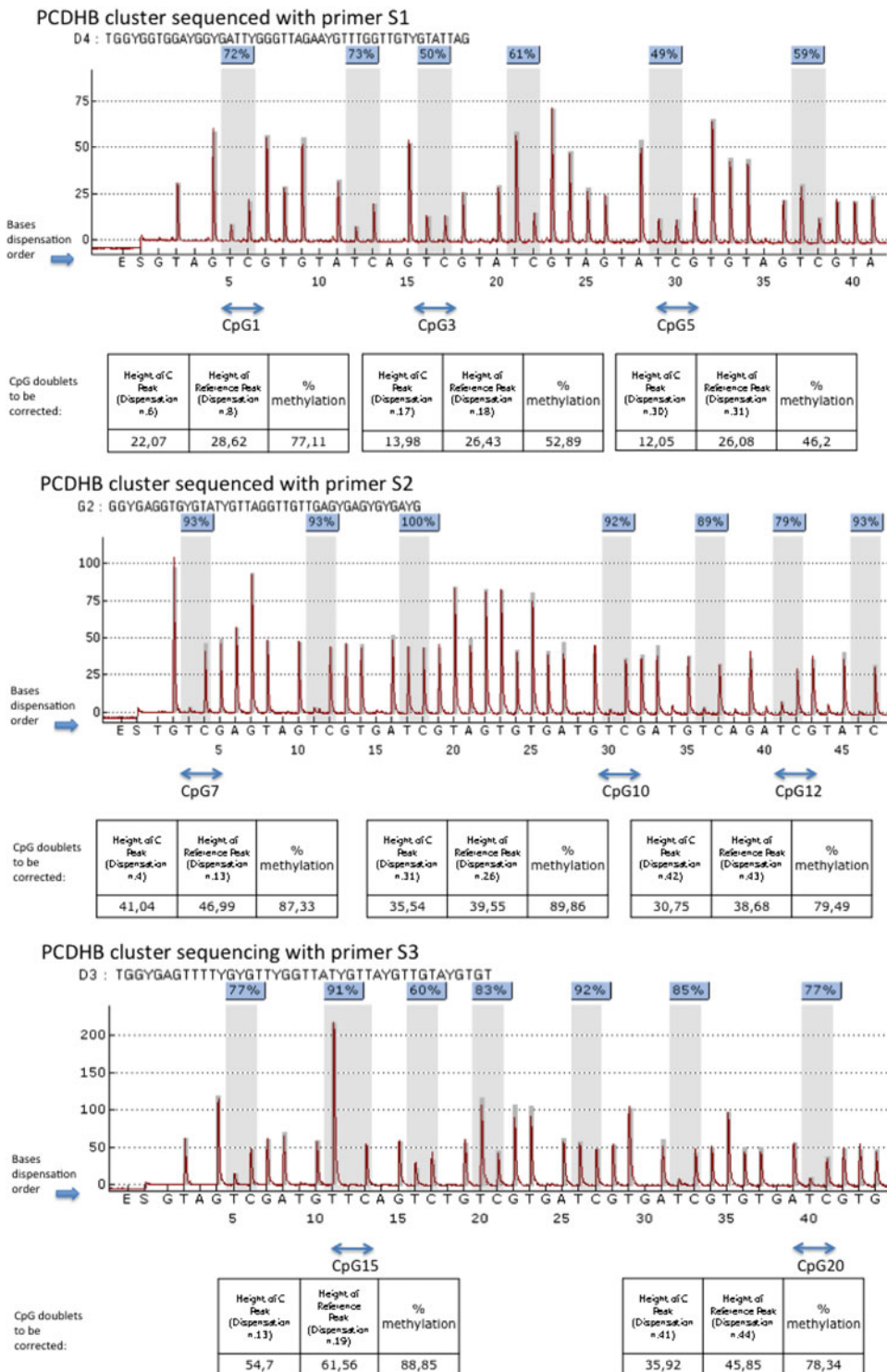


Fig. 5 Representative Pyrograms with the three sequencing primers (S1, S2, S3). Below each Pyrogram the calculation of the correct methylation value utilizing the relative reference peaks as indicated in Table 2 is shown

2. The introduction of a degenerate base could potentially introduce a bias in the PCR amplification, influencing the subsequent Pyrosequencing reaction [4]. To rule out a preferential amplification, the PCR products resulting from an annealing temperature gradient were analyzed by Pyrosequencing and no differences in the level of methylation were found.
3. If the amount of modified DNA is limited, *see* ref. [5].
4. The Peaks Heights function is a possible output of the Pyro Q-CpG software that can be activated from the “Reports” menu, a pull-down menu at the top of the User Interface of the software. For each nucleotide dispensation set beforehand, the height of the peaks of light is recorded in an Excel® file.

References

1. Abe M, Ohira M, Kaneda A, Yagi Y et al (2005) CpG island methylator phenotype is a strong determinant of poor prognosis in neuroblastomas. *Cancer Res* 65:828–834
2. Banelli B, Brigati C, Di Vinci A et al (2012) A pyrosequencing assay for the quantitative methylation analysis of the PCDHB gene cluster, the major factor in neuroblastoma methylator phenotype. *Lab Invest* 92:458–465
3. Banelli B, Merlo DF, Allemanni G et al (2013) Clinical potentials of methylator phenotype in stage 4 high-risk neuroblastoma: an open challenge. *PLoS One* 8: e63253
4. Shen L, Guo Y, Chen X et al (2007) Optimizing annealing temperature overcomes bias in bisulfite PCR methylation analysis. *Biotechniques* 42:48–58
5. Tost J, El abdalaoui H, Gut IG (2006) Serial pyrosequencing for quantitative DNA methylation analysis. *Biotechniques* 40: 721–726

Chapter 15

Assessment of Changes in Global DNA Methylation Levels by Pyrosequencing® of Repetitive Elements

Ali M. Tabish, Andrea A. Baccarelli, Lode Godderis,
Timothy M. Barrow, Peter Hoet, and Hyang-Min Byun

Abstract

Transposable elements (TE) comprise half of the human genome. LINE-1 and *ALU* are the most common TE, and they have been used to assess changes in the DNA methylation of repetitive elements in response to intrinsic and extrinsic cellular events. Pyrosequencing® is a real-time sequencing technology that enables quantitative assessment of TE methylation at single-base resolution. In Pyrosequencing, a region of interest is first amplified from bisulfite-converted DNA by polymerase chain reaction (PCR), before PCR amplicons are rendered single stranded and annealed with the Pyrosequencing primer prior to sequencing. In this chapter, we provide an overview of the analysis of repetitive element DNA methylation by bisulfite Pyrosequencing, and we describe a protocol that can be used for such purposes.

Key words Repetitive elements, LINE-1, *Alu*, Epigenetics, Global DNA methylation, Bisulphite pyrosequencing, CpG-dinucleotides

1 Introduction

Transposable elements (TEs) comprise approximately half of the human genome [1]. Among TEs, the long interspersed elements (LINEs) and short interspersed elements (SINEs) are most common and well-studied. The most common SINEs in mammals are the *Alu* elements. LINE-1 and *Alu* methylation can be altered in response to stress, infection, diseases [2], and they are linked with the genomic instability that is implicated in genetic disorders such as gastrointestinal stromal tumors, myeloma, ependymomas, and lung cancer [3]. The methylation of these elements reduces genomic instability by controlling LINE-1 and *Alu* induced retrotransposition. Different classes of TEs, such as retrotransposons and DNA transposons, display different susceptibility to changes in DNA methylation in response to environmental exposures or disease [4, 5]. Furthermore, different subfamilies of LINE-1 and *Alu* elements, generated through mutations over

the evolutionary history of the elements, show differential sensitivity to the exposome [2].

Bisulfite Pyrosequencing[®] is a technique based upon the “sequencing by synthesis” approach. It provides quantitative and highly reproducible methylation data at single-base resolution, and it requires relatively low quantities of DNA [6]. Pyrosequencing assays for the analysis of TE (LINE or *Alu*) DNA methylation interrogate several thousand copies of elements across the genome, as opposed to one unique locus. CpG sites are particularly enriched at TEs, accounting for >65 % of those found throughout the human genome [7], and therefore the methylation of these elements has been suggested as an indicator of global DNA methylation [8–10]. However, several investigations have reported poor correlations in normal tissues between the DNA methylation of TEs and global genomic DNA methylation, as determined by the “gold standard” of high performance liquid chromatography (HPLC) [11, 12]. Nonetheless, LINE-1 and *Alu* methylation can serve as a robust surrogate marker of global DNA methylation for some purposes, such as in cancer cells, where profound TE demethylation contributes to a global loss of nuclear DNA methylation [13, 14].

Here, we describe a protocol routinely used in our lab for the analysis of DNA methylation of human TE (*see* **Note 1**).

2 Materials

1. Human genomic DNA: up to 1.0 µg of DNA is required (*see* **Note 2**).
2. Bisulfite-converted DNA: this should be prepared using 0.25–1.0 µg of isolated genomic DNA in conjunction with a commercially available bisulfite conversion kit. The eluted bisulfite-converted DNA (20–40 µL) can be aliquoted and stored at –80 °C until required for use.
3. LINE-1 PCR primer sets (100 µM): for the PCR-based amplification of LINE-1 subfamilies from bisulfite-converted DNA and the performance of bisulfite pyrosequencing (*see* **Note 3** and Table 1).
4. Streptavidin Sepharose[®] High Performance beads (GE Healthcare).
5. Pyrosequencer and Vacuum Workstation, e.g., PyroMark[®] Q96 MD system (QIAGEN[®]).
6. PSQ HS 96 sample prep thermo plate kit (QIAGEN).
7. PyroMark Q96 HS Plate (QIAGEN).
8. PyroMark[®] Gold Q96 Reagents: enzymes, substrate, and nucleotides required to perform the pyrosequencing reaction (QIAGEN).
9. PyroMark Binding Buffer, Annealing Buffer, Denaturation Solution, and Wash buffer (QIAGEN).

Table 1
Bisulfite Pyrosequencing assays for human LINE-1 subfamilies

LINE-1 subfamilies	Forward primer	Reverse biotinylated primer	Sequencing primer	Sequence to analyze	Annealing temperature (°C)	Amplicon size (bp)
L1PA5	TTAGTTAAGGG AAGAGGGGATAAA	ATAAACATAAAACCC TCTAAACCAACA	TTAGTTAAGGAAAGA	GGGGATAAAC/ TGGTAATTTGAAAAATC	40	140
L1PA2	TTAGATAGTGGYG TAGGTTAGTGGGT	CCTCCRAACCAAATA TAAATATAATCT	GAGTTAAAGAAAGGG	GTGAC/TGGAC/ TGTATTTGGAAAAATC/ TGGGTTAATTTTAT	55	232
L1Hs	TTTTGAGTTAGT GTGGGATATA	AAATCAAAAAAATT CCCTTTC	AGTTAGGTGT GGATATAGT	TTC/TGTGGTGC/TGTC/ TGTTTTTTAAGTC/ TGGTTTGAAAAAGC/TGTA	56.3	156
L1Ta	GGGTTAGGGAG TTTTTTTT	CTCTAAACCAAATA TA AAATATA	GGGTTAGGGA GTTTTTTTT	C/TGAGTTAAAGAAAGGGGTGAC/ TGGAC/TGTATTTGAAAAATC/ TGGGTTAATTTTAT	55	145

3 Method

3.1 PCR Amplification of Bisulfite-Converted DNA

PCR-based amplification of the TE to be analyzed can be performed using 15 μL of GoTaq Green Master mix (Promega), 10 pmol forward and 10 pmol reverse primers, 1.5 mM MgCl_2 , 10–20 ng of bisulfite-treated DNA, and water to reach a final volume of 30 μL . The reaction can then be performed under the following thermocycling conditions: 15 min at 95 °C; 45 cycles of (94 °C for 30 s; annealing temperature (Table 1) for 30 s; 72 °C for 30 s); 72 °C for 5 min. The specificity and yield should be verified for each reaction by visualization of PCR amplicons run on a 2 % agarose gel under electric field (100 V) for 20 min. Amplicon sizes for published assays interrogating different members of LINE-1 sub-families are given in Table 1 (*see* Note 4).

3.2 Bisulfite Pyrosequencing

3.2.1 Sample Preparation

1. Prepare the PyroMark Binding Buffer mixture as follows: 38 μL PyroMark Binding Buffer, 2 μL streptavidin Sepharose[®] beads, 30 μL dH_2O .
2. Dispense 70 μL of Binding Buffer mixture into each well of a 96-well plate and add 10 μL of PCR product.
3. After sealing the plate with an adhesive cover, incubate the plate for 10 min with vigorous shaking.
4. Prepare the Annealing Buffer mixture as follows: 11.64 μL PyroMark annealing buffer, 0.36 μL sequencing primer (0.3 μM concentration).
5. Dispense 12 μL of Annealing Buffer mixture into a PyroMark Q96 HS Plate.

3.3 Strand Separation

6. Fill the Vacuum Workstation with the corresponding buffers (~180 mL of 70 % ethanol, Denaturation Solution, Wash Buffer, and Milli-Q grade water).
7. Apply vacuum to the vacuum prep tool.
8. Prime the probes of the vacuum prep tool by immersion in the Milli-Q water for approximately 20 s.
9. Capture the beads containing the immobilized PCR templates on the filter-probes by slowly lowering the vacuum prep tool into the 96-well PCR plate from step 3.
10. Place the vacuum prep tool into 70 % ethanol and let the solution flush through the filters for 5 s.
11. Place the vacuum prep tool into the Denaturation Solution and flush it through the filters for 5 s.
12. Place the vacuum prep tool into the Wash Buffer and flush it through the filters for 5 s.

13. Switch off the vacuum using the switch on the vacuum prep tool handle and completely remove it from the vacuum station.
14. Release the beads in a PyroMark Q96 HS Plate filled with the annealing buffer mixture by gently shaking the vacuum prep tool.
15. Seal the PyroMark Q96 HS Plate with an adhesive cover.

3.4 Primer Annealing

16. Heat the PyroMark Q96 HS Plate at 80 °C for 2 min using the PSQ HS 96 sample prep thermo plate kit.
17. Remove the plate from the heating block and leave the samples on the bench for 10 min in order to cool to room temperature.

3.5 Pyrosequencing Reaction

18. Load the Pyrosequencing reagents into the cartridges. The volumes required for each reagent are automatically calculated and described within the PyroMark CpG software.
19. After performing the dispensing tip test, run the assay within the PyroMark CpG software (*see Note 5*). Once the run is completed, the data can be extracted (*Fig. 1*).

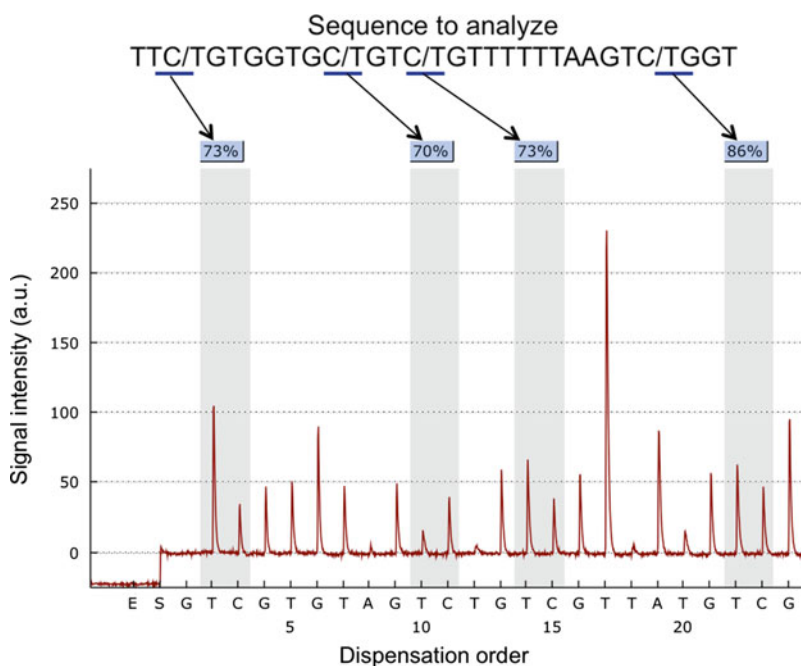


Fig. 1 Pyrogram® of multiple CpG sites within the LINE-1 sequence. The Y-axis gives intensity (in a.u.; arbitrary units), and the X-axis is the nucleotide dispensation order. The LINE-1 sequence analyzed in this assay is given at the *top* of the figure. Methylation at four CpG sites (highlighted in *gray*) is quantified in a single Pyrosequencing run. Above the *gray bars* are the percentage methylation levels at each CpG site

4 Notes

1. Repetitive elements DNA methylation in other species, such as mice, rat and monkey, can be measured using bisulfite Pyrosequencing. However, the repetitive element sequences are different from those in humans, and therefore the bisulfite pyrosequencing primers have to be carefully designed for each species. These primer sequences are available upon request.
2. DNA methylation analysis using bisulfite pyrosequencing is applicable to a wide variety of biological samples, such as DNA isolated from cultured cells, whole blood samples and other body fluids (buccal cells, saliva, nasal swab samples, etc.), tumor tissue from different sources, and processed samples such as formalin-fixed paraffin embedded tissue samples.
3. The conventional LINE-1 assay [8] is based upon the L1HS sequence (Table 1). The “LINE-1 assay” referred to throughout this chapter is that based upon interrogation of L1HS sequences.
4. Appropriate quality controls are required at each stage of the experiment. Quality control samples should be run alongside the samples of interest for the bisulfite-conversion of DNA, the PCR reaction, and the pyrosequencing run.
5. The PyroMark CpG software has in-built quality control parameters to assess the overall performance of the sequencing run. The software also supports analysis, such as methylation frequency, mean methylation values per well, and replicates and deviation from expected methylation pattern. One important aspect of methylation analysis by Pyrosequencing is to assess the overall bisulfite conversion efficiency. Since all non-CpG cytosine residues are converted to uracil and then to thymine in the subsequent PCR reaction, the relative incorporation of thymines and cytosines into the elongated DNA at these sites informs upon the efficiency of the bisulfite conversion reaction.

References

1. Lander ES, Linton LM, Birren B, International Human Genome Sequencing C et al (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921
2. Byun HM, Motta V, Panni T et al (2013) Evolutionary age of repetitive element subfamilies and sensitivity of DNA methylation to airborne pollutants. *Parti Fibre Toxicol* 10:28
3. Su J, Shao X, Liu H et al (2012) Genome-wide dynamic changes of DNA methylation of repetitive elements in human embryonic stem cells and fetal fibroblasts. *Genomics* 99:10–17
4. Choi SH, Worswick S, Byun HM et al (2009) Changes in DNA methylation of tandem DNA repeats are different from interspersed repeats in cancer. *Int J Cancer* 125:723–729
5. Guo L, Byun HM, Zhong J et al (2014) Effects of short-term exposure to inhalable particulate matter on DNA methylation of tandem repeats. *Environ Mol Mutagen* 55:322–335

6. Florea A-M (2013) DNA methylation pyrosequencing assay is applicable for the assessment of epigenetic active environmental or clinical relevant chemicals. *Biomed Res Int* 2013: 486072
7. Lai RK, Chen Y, Guan X et al (2014) Genome-wide methylation analyses in glioblastoma multiforme. *PloS One* 9:e89376
8. Yang AS, Estecio MR, Doshi K et al (2004) A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. *Nucleic Acids Res* 32:e38
9. Baccarelli A, Wright R, Bollati V et al (2010) Ischemic heart disease and stroke in relation to blood DNA methylation. *Epidemiol* 21: 819–828
10. Zhu ZZ, Hou L, Bollati V et al (2012) Predictors of global methylation levels in blood DNA of healthy subjects: a combined analysis. *Int J Epidemiol* 41:126–139
11. Wagner I, Capesius I (1981) Determination of 5-methylcytosine from plant DNA by high-performance liquid chromatography. *Biochim Biophys Acta* 654:52–56
12. Feinberg AP, Vogelstein B (1983) Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 301:89–92
13. Baba Y, Huttenhower C, Nosho K et al (2010) Epigenomic diversity of colorectal cancer indicated by LINE-1 methylation in a database of 869 tumors. *Mol Cancer* 9:125
14. Lisanti S, Omar WA, Tomaszewski B et al (2013) Comparison of methods for quantification of global DNA methylation in human cells and tissues. *PloS One* 8:e79044

Free ebooks ==> www.Ebook777.com

Chapter 16

Global Analysis of DNA 5-Methylcytosine Using the Luminometric Methylation Assay, LUMA

Karin Luttropp, Louise K. Sjöholm, and Tomas J. Ekström

Abstract

The study of epigenetic alterations of the genome is becoming increasingly important in order to understand how environment and genetic background interact to build and regulate the functional genome. There are several types of epigenetic modifications to both DNA and histone proteins in eukaryotic cells; chiefly studied among these are changes to cytosine, where methylation of the 5-carbon position is the most prominent. Although this has many consequences for gene regulation and cell differentiation, other modifications have recently emerged as biologically relevant. Since global DNA methylation states may be used as a general measure of the methylome, cost-effective, rapid, and specific analytical tools are wanted.

This protocol described here focuses on the Luminometric Methylation Assay (LUMA), a method which analyzes global DNA 5-methylcytosine (5mC) through the use of restriction enzymes and detection with Pyrosequencing®. Up to 96 samples can be simultaneously analyzed. In contrast to the majority of other methods focused on 5mC analysis, with appropriate enzymes, LUMA does not appear to detect 5-hydroxymethylcytosine (5hmC) and is therefore more specific than most 5mC techniques.

Key words LUMA, Global DNA methylation, 5-Methylcytosine, Epigenetics, Pyrosequencing

1 Introduction

1.1 Background

The measurement of epigenetic changes is rapidly becoming an integral part of any genetic analysis aiming to determine effects of inter-individual genetic changes, differences in gene expression or the impact of environmental factors on the genome, to name a few scientific objectives. Traditionally, the primary focus has been on 5-CpG methylation, and several methods aiming at measuring this particular type of epigenetic alteration exist [1]. However, most of them suffer from several drawbacks such as low throughput, utilization of radioactive isotopes, high cost, and/or the need for large amounts of DNA. The use of restriction endonucleases was early recognized as a way to measure global DNA CpG methylation [2]. This strategy employs isoschizomer endonucleases, which have the

same recognition sequence, where one enzyme is sensitive to CpG methylation whereas the other is not. Frequently, *HpaII* and *MspI* are used; both have the recognition sequence CCGG, but while *MspI* will cut the recognition sequence regardless of methylation status, *HpaII* will not perform any restriction on the internal cytosine if methylated (C^mCGG) [2]. In human cells, CCGG sequences in the genome are usually methylated to 50–70 % (*see Note 1*). There are several methods which utilize *HpaII* and *MspI* to study DNA methylation; methylation-sensitive arbitrarily primed polymerase chain reaction (AP-PCR) [3], self-primed in situ labeling (SPRINS) [4], differential methylation hybridization (DMH) [5], non-methylated genomic sites coincidence cloning (NGSCC) [6], methylation target array (MTA) [7], and reduced representation bisulfite sequencing (RRBS) [8], to name a few. Reference [9] contains a more extensive overview of techniques for DNA methylation analysis.

LUMA is a development of previously established methods. In the “cytosine extension assay” developed by Pogribny et al., *HpaII* and *MspI* endonuclease restriction is combined with a single nucleotide extension, in which radioactively labelled [3H]dCTP is incorporated. Thus, the amount of incorporated [3H]dCTP is inversely correlated with DNA methylation [10]. Following the work of Pogribny et al., a version using biotinylated dCTP was developed [11], in which the quantity of DNA methylation is defined as the *HpaII*/*MspI* ratio. When the DNA is completely unmethylated, the amount of *HpaII* restriction equals that of *MspI* restriction, and the ratio would be 1.0. Conversely, the ratio would approach 0.0 in the case of full DNA methylation.

In this chapter, we describe the Luminometric Methylation Assay (LUMA), which replaces the use of radioactive single nucleotides by a Pyrosequencing[®] reaction [12]. Several studies have successfully applied LUMA for DNA methylation analysis [13–27].

1.2 Principle of LUMA

The LUMA method usually utilizes the isoschizomers *HpaII*/*MspI* (or some other methylation-sensitive/insensitive isoschizomer enzyme leaving a 5'-overhang) to perform enzymatic restriction of DNA. The enzymatically treated DNA is then analyzed by a luminometric polymerase extension assay to quantify the amount of restriction cleavage by each of the enzymes. The relative amount of DNA methylation is then expressed as an *HpaII*/*MspI* ratio. Two parallel reactions, one with *HpaII* and one with *MspI*, are run. One great advantage of LUMA is the ability to enable normalization between runs and for DNA input by including *EcoRI* in all reactions. *EcoRI* has the recognition sequence GAATTC, and is thus unaffected by CpG methylation (*see Note 2*). After *HpaII* or *MspI* restriction, there is a resulting 5'-CG overhang, whereas *EcoRI* restriction yields a 5'-AATT overhang. Using the Pyrosequencing platform, nucleotides are added stepwise in a predetermined order,

discriminating the CG from the AATT overhangs. As the nucleotides (dNTP) are incorporated with DNA polymerase, inorganic pyrophosphate (PPi) is released and converted to ATP by ATP-sulfurylase and adenosine-5'-phosphosulfate. The resulting ATP is then utilized by luciferase to convert luciferin to oxyluciferin by luciferase, which produces visible light. The light is proportional to the amount of dNTP incorporation in the original 5'-overhang produced by the restriction enzyme. The light is then detected by a charge coupled device (CCD) camera and visualized as peaks in the software [28]. For LUMA, dNTPs are added in four sequential steps (Step 1: dATP α S (*see* **Note 3**), Step 2: dGTP + dCTP, Step 3: dTTP and Step 4: dGTP + dCTP). Peaks following dATP α S (Step 1) and dTTP (Step 3) dispensations both correspond to the amount of *Eco*RI restriction (which leaves AATT overhangs) and are expected to be equal to one another since the amount of dATP and dTTP incorporation should be the same. dCTP and dGTP are added simultaneously in dispensation 2, and the corresponding peak represents the amount of *Hpa*II or *Msp*I restriction. Finally, dCTP and dGTP are added again in step 4 as a control that all *Hpa*II or *Msp*I overhangs were filled in during step 2. The peak height following step 4 should be zero or close to zero. Figure 1 illustrates the schematic principle of LUMA.

1.3 Interesting Considerations

The analysis of the DNA 5mC modification has recently become more complicated when it was realized that hydroxylation of 5mC to 5-hydroxymethylcytosine (5hmC) is more common than previously known, and furthermore, that this modification has biological significance [29]. All methods employing bisulfite conversion to discriminate between cytosine and 5mC do not discriminate between 5mC and 5hmC. Although very recent methods are now available that can remedy this, a substantial amount of data in the scientific literature are not entirely correct [30]. The LUMA method usually relies on restriction enzymes *Hpa*II and *Msp*I (*see* **Note 4**). A recent article reported that these enzymes are in fact sensitive to 5hmC [31]. Although we have not investigated this issue, the consequence of this sensitivity means that LUMA, employing *Hpa*II/*Msp*I, is in fact discriminating between global 5mC and 5hmC, and only 5mC in CCGG sites will be assessed.

2 Materials

2.1 Assessment of DNA Quality

1. 10 $\times$ TBE buffer: 0.89 M Tris-Borate, pH 8.3, 20 mM Na₂EDTA.
2. Standard agarose (1 %).
3. Suitable DNA ladder (e.g., 1 kb) to be visualized on the agarose gel.

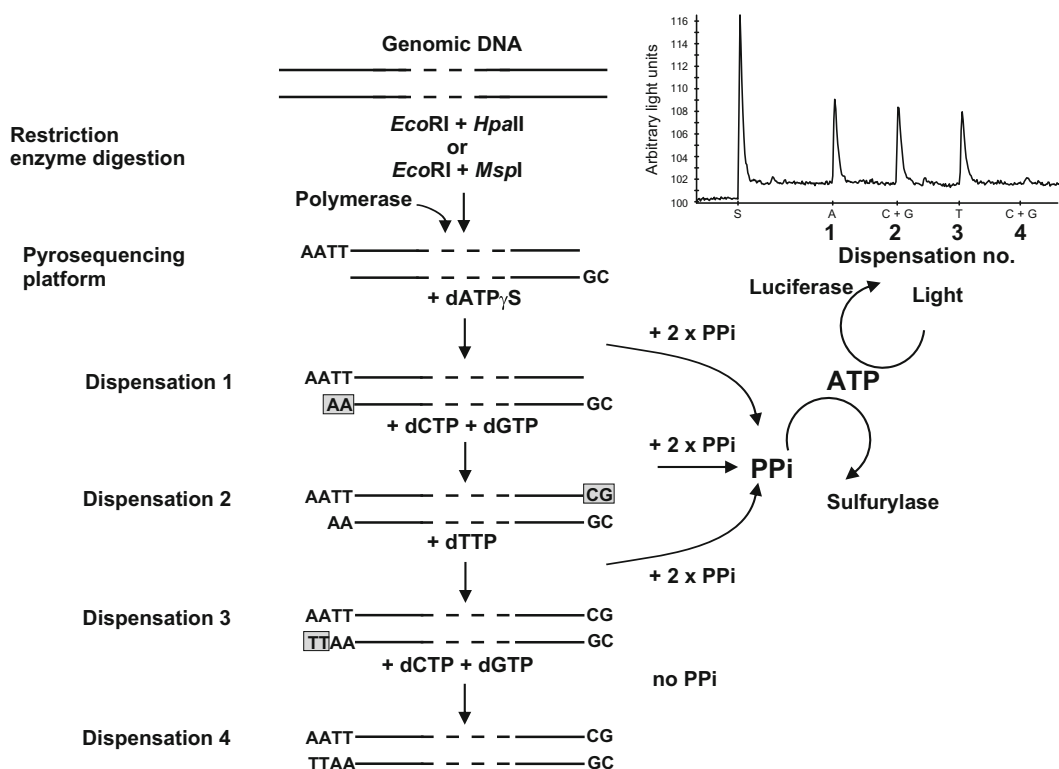


Fig. 1 Illustration of the LUMA assay. Restriction enzymes (usually either *HpaII* + *EcoRI* or *MspI* + *EcoRI*) digest the genomic DNA, after which the degree of enzymatic restriction is quantified by a polymerase extension assay based on a four-step Pyrosequencing[®] reaction. Inorganic pyrophosphate (PPi) is released after each nucleotide incorporation, which is consequently converted to ATP by ATP-sulfurylase. The resulting ATP is then used by luciferase to activate luciferin. The amount of luciferin activation produces a proportional amount of visible light, which is detected by a CCD camera. Thus, the number of overhangs produced by the respective restriction enzymes corresponds to the amount of light produced, and the height of the light peaks are used for this quantification

4. GelRed (or equivalent) dissolved in water (Biotium) (*see Note 5*).
5. UV-table for visualizing agarose gels.
6. NanoDrop for determining DNA concentration and purity.

2.2 Enzyme Restriction of Genomic DNA

1. Restriction enzyme *HpaII*, 10 U/ μ L (New England Biolabs or other provider).
2. Restriction enzyme *MspI*, 20 U/ μ L (New England Biolabs or other provider).
3. Restriction enzyme *EcoRI*, 20 U/ μ L (New England Biolabs or other provider).
4. Tango[™] buffer 10 $\times$: 33 mM Tris-Acetate, pH 7.9, 10 mM Mg-Acetate, 66 mM K-Acetate, 0.1 mg/mL BSA) (Fermentas).
5. DNase-free water.
6. Pyrosequencing plates (Qiagen).

2.3 Pyrosequencing Analysis

1. Pyrosequencing instrument. Other instruments can optionally be used (*see* **Notes 6** and **7**).
2. PyroMark® Annealing buffer for Pyrosequencing (Qiagen).
3. Dispensation cartridge for Pyrosequencing instrument (Qiagen).
4. Pyrosequencing Gold Kit (Qiagen).

3 Methods

The LUMA workflow is illustrated in Fig. 2.

LUMA is normally performed using *HpaII*/*MspI*. However, other methylation-sensitive restriction enzymes can also be used (*see* **Note 4**). The method described below can be performed on all different Pyrosequencing platforms (*see* **Notes 6** and **7**). For convenience, three Stop points have been indicated, where the procedure can be paused and samples stored at -20°C .

3.1 Assessment of DNA Quality

1. Since LUMA relies on enzymatically created cuts in the genome, it is vital that the DNA is of high quality. To check the integrity of DNA, it is highly recommended that the samples are run on a 1 % agarose gel with GelRed (or equivalent) staining prior to LUMA analysis (*see* **Note 5**). DNA of sufficient

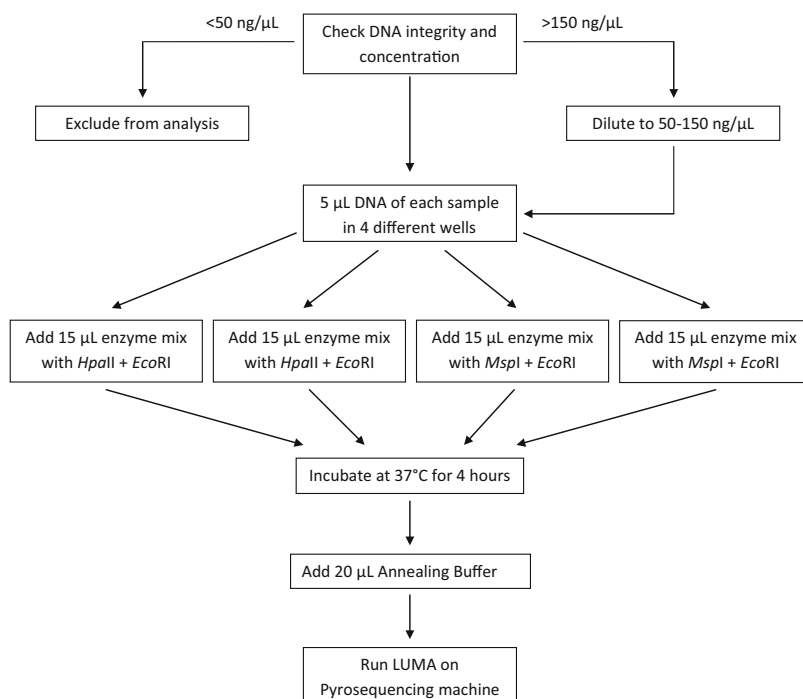


Fig. 2 Flowchart of the LUMA process

quality will appear as a strong band with high molecular weight, without any visible smear below the band.

2. To ensure that the DNA is pure and that the concentration is sufficient, a spectrophotometric measurement is used. DNA samples with protein and/or RNA contamination should not be included in the LUMA analysis, and any samples with a concentration below 50 ng/μL should be avoided.

Stop point 1: The samples can be frozen at –20 °C and used later.

3.2 Preparation of Genomic DNA for LUMA

1. DNA samples that have been tested according to Subheading 3.1 are prepared for LUMA analysis. The preferred concentration is between 50 and 150 ng/μL, but it is not necessary that all samples have equal concentrations since the *EcoRI* internal normalization control will compensate for differences (*see steps 3–5* in Subheading 3.5). Samples should be diluted in DNase-free water.
2. 5 μL of each diluted sample (at a concentration of 50–150 ng/μL) is dispensed onto Pyrosequencing plates for enzymatic restriction. It is highly recommended that all samples are performed in duplicate throughout the procedure. Also, keep in mind that all samples are present in two reactions for the enzyme restriction—one for the *HpaII* restriction and one for the *MspI* restriction. Including the recommended duplicates, this means that each sample is dispensed in four wells, amounting to a required total volume of 20 μL for each sample.

Stop point 2: The samples can be used for LUMA immediately, or the plates can be sealed and stored at –20 °C.

3.3 Enzyme Restriction of Genomic DNA

Master mixes containing enzymes are prepared for restriction digestion of genomic DNA. One mix (Mix A) contains *HpaII*, *EcoRI*, Tango buffer and water, whereas the other (Mix B) has the same ingredients except that *HpaII* is substituted with *MspI*. Both mixes are calculated based on a total volume of 20 μL including 5 μL DNA. Master mixes specified below can be scaled up to an appropriate number of samples.

1. Mix A is prepared in the following way (volumes per sample):
 - 13 μL DNase-free water.
 - 2 μL 10× Tango™ buffer.
 - 5 U *EcoRI* (0.25 μL).
 - 5 U *HpaII* (0.5 μL).
2. Mix B is prepared in the following way (volumes per sample):
 - 13 μL DNase-free water.
 - 2 μL 10× Tango™ buffer.

5 U *Eco*RI (0.25 μ L).

5 U *Msp*I (0.25 μ L).

3. Add 15 μ L of either Mix A or Mix B to the 5 μ L DNA sample on a Pyrosequencing plate. Mix by pipetting up and down, seal the plates and incubate at 37 °C for 4 h. No heat inactivation is performed after the incubation.

Stop point 3: The reactions can be frozen at this step, preferably at -20 °C.

3.4 Pyrosequencing Assay

1. Program the instrument by setting up a run in “SNP” mode. The sequence to analyze should be set as AC/TCGA, which results in an ACTCGA dispensation order for the nucleotides (*see Note 8*).
2. Add 20 μ L of Pyrosequencing Annealing Buffer to each reaction. This should be done as soon as possible following the enzyme incubation as described in Subheading 3.3.
3. Prepare the nucleotides used in the Pyrosequencing reaction. The volumes specified below are for 96 samples (1 plate):

Dilute 50 μ L dATP α S with 50 μ L ddH₂O.

Dilute 50 μ L dTTP with 50 μ L ddH₂O.

Mix 50 μ L dCTP and 50 μ L dGTP.

Add the nucleotides to the Pyrosequencing cartridge as shown in Fig. 3.

4. Reconstitute the enzyme and substrate mix from the Pyro Gold kit by adding 620 μ L ddH₂O to each of the vials. One vial is enough for 96 samples. Add the enzyme mix to the “E” compartment of the Pyrosequencing cartridge, and the substrate mix to the “S” compartment, as shown in Fig. 3.
5. Start the run after inserting the cartridge and the plate in the instrument. A typical run is displayed in Fig. 4.

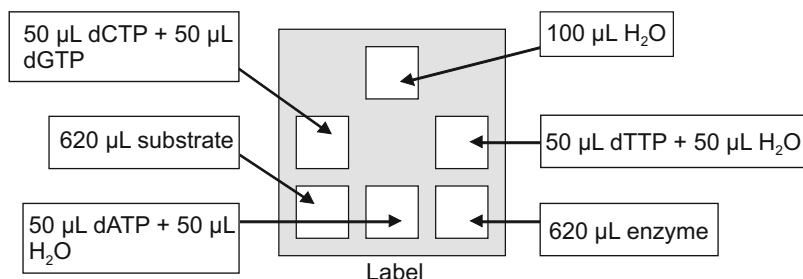


Fig. 3 Addition of reagents to the Pyrosequencing cartridge for LUMA. Note that the label on the cartridge should face front. The volumes given will be sufficient for one 96-well plate

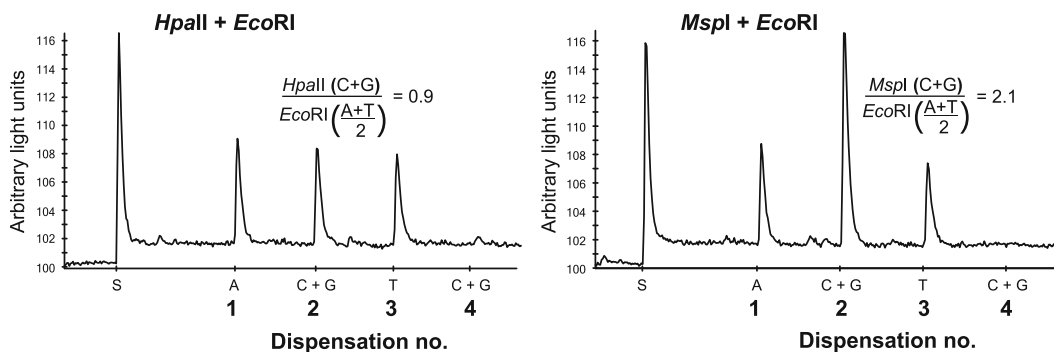


Fig. 4 Typical LUMA results for a sample of human genomic DNA analyzed by a PSQ96 MA Pyrosequencer. The graphs show the luminometric output from two representative LUMA runs of human lymphocyte DNA, using *HpaII* + *EcoRI*, and *MspI* + *EcoRI*. The A and T peaks denote the additions of dATP α S and dTTP, respectively, which extend all overhangs generated by *EcoRI*. The second peak denotes the second nucleotide dispensation and is an indication of the degree of *HpaII* or *MspI* restriction. The first peak, designated “S,” is the peak resulting from addition of substrate mix to the enzyme mix, and is always present at the starting point of Pyrosequencing reactions

3.5 Data Analysis

1. When the Pyrosequencing run has terminated, open the resulting data file and select “AQ mode” for the data analysis. After the analysis has finished, export the peak height data from the “Peak height” menu.
2. Open the exported data in Excel. The peak heights for all six nucleotide dispensations (numbered as 1–6 in the order of dispensation) are specified (*see Note 9*). First, perform a quality check of your data. The first three peaks, following the substrate peak, should be substantially higher than the last three (CGA), which should be very small and preferably non-existent. If these last peaks are high, especially the second C and A dispensations, this usually indicates fragmented DNA (*see Note 9*).
3. To obtain the *HpaII*/*MspI* ratio as an indication of degree of methylation, do as follows: Calculate the average of peak 1 and peak 3 to get an average peak height for *EcoRI* restriction (*see Note 10*).
4. Calculate the ratio of peak 2 to the averaged *EcoRI* peak for *HpaII* and *MspI* digestions separately, to get normalized values for both enzymes (*see Note 11*).
5. Calculate the ratio of normalized *HpaII* to normalized *MspI* to get the *HpaII*/*MspI*-ratio. This value indicates the degree of methylation; the higher the ratio, the lower the amount of methylation (since the *HpaII* digestion increases with decreasing methylation levels; *see Note 12*).

4 Notes

1. *HpaII* recognition sites are distributed throughout the genome. Although *HpaII* sites are enriched 15-fold in CpG islands [32], these sites represent only 12 % of the total *HpaII* sites in the entire human genome [33]. Therefore, *HpaII*/*MspI* ratios can be considered representative for whole genome methylation.
2. *EcoRI* cutting may be impaired if one or both of the 3'-C in the recognition sequence (GAATTC) are methylated (<http://rebase.neb.com/cgi-bin/msget?EcoRI+5>). Please see **Note 11** for more information on how this affects data analysis.
3. In Pyrosequencing, the dATP nucleotide has been chemically modified to dATP α S. This enables it to become incorporated in the growing DNA strand and yield a light signal, but prevents it from interacting with the ATP-dependent enzyme luciferase. However, the dATP α S is still a weak substrate for luciferase, which results in slightly higher A peaks than T peaks, but this is normal and not a source of error.
4. While LUMA as described in this article is based on the isoschizomers *HpaII* and *MspI*, this will not provide full coverage of CpG sites in the genome since these enzymes only recognize the CCGG sequence. However, LUMA allows the use of other methylation-sensitive restriction enzymes, given that they yield suitable 5'-overhangs. Thus, using enzymes with another recognition sequence in addition to *HpaII* and *MspI* could provide an improved coverage of CpG or CpNpG sites. For example, cytosine methylation in CCWGG sequences has been analyzed using the restriction enzymes *Psp6I* and *AjnI* [13].
5. GelRed is generally preferred to EtBr as the latter is carcinogenic, whereas the former is reported to be less toxic.
6. Pyrosequencing instruments may not be available to all research laboratories, but it is possible to apply LUMA using other platforms. Preliminary data in our hands indicate that a 96-well luminometer equipped with an automatic dispenser can substitute the Pyrosequencing platform. A Luminoskan Ascent microplate Luminometer from Thermo Scientific may replace a Pyrosequencer instrument for running LUMA.
7. Using a high sensitivity Pyrosequencing instrument, less DNA (200 ng) may be used for the analysis.
8. The Pyrosequencing software requires the position of a SNP to be specified in the dispensation order when running the machine in SNP mode. Therefore, the dispensation sequence includes a "fake" SNP, which does not have any effect on the dispensation of the nucleotides.

9. It is not uncommon to see small peaks among the last three (CGA) nucleotide dispensation; this is especially true for peak 6 due to the chemical modification of dATP α S. Minor signals can be tolerated, as long as they are substantially smaller than peaks 1–3; they should be no more than approximately 10 % of the peak height seen in peaks 1–3.
10. While it is normal for peak 1 to be somewhat higher than peak 3 due to the chemically modified dATP α S, these two peaks should not differ too much.
11. The theoretical ratio of *MspI*/*EcoRI* is 2.8–3.0 in mammalian genomes. Therefore, the *MspI*/*EcoRI* ratio should normally not exceed 3, as there are no more than three times as many *MspI* restriction sites as there are *EcoRI* sites. A ratio >3 may indicate that the restriction has failed for one or both of the enzymes, or that the DNA was degraded. However, *MspI*/*EcoRI* ratios may be >3 since *EcoRI* restriction can be affected by methylation (please see **Note 2** for more information). Since *EcoRI* is added as normalizer to both reaction tubes from the same sample, this is normally not an issue.
12. It is recommended that the percentage of methylation is calculated based on the LUMA results with the following formula: $\text{Methylation \%} = 100 \left[1 - \left(\text{HpaII} / \text{EcoRI} / \text{MspI} / \text{EcoRI} \right) \right]$. This will provide the percentage of methylation and a clearer and logical data presentation.

References

1. Fraga MF, Esteller M (2002) DNA methylation: a profile of methods and applications. *Biotechniques* 33:632–649
2. Cedar H, Solage A, Glaser G et al (1979) Direct detection of methylated cytosine in DNA by use of the restriction enzyme *MspI*. *Nucleic Acids Res* 6:2125–2132
3. Liang G, Gonzalgo ML, Salem C et al (2002) Identification of DNA methylation differences during tumorigenesis by methylation-sensitive arbitrarily primed polymerase chain reaction. *Methods* 27:150–155
4. Andersen CL, Koch J, Kjeldsen E (1998) CpG islands detected by self-primed in situ labeling (SPRINS). *Chromosoma* 107:260–266
5. Yan PS, Wei SH, and Huang TH (2002) Differential methylation hybridization using CpG island arrays. *Methods Mol Biol* 200:87–100
6. Azhikina T, Gainetdinov I, Skvortsova Y et al (2004) Non-methylated genomic sites coincidence cloning (NGSCC): an approach to large scale analysis of hypomethylated CpG patterns at predetermined genomic loci. *Mol Genet Genomics* 271:22–32
7. Chen CM, Chen HL, Hsiao TH et al (2003) Methylation target array for rapid analysis of CpG island hypermethylation in multiple tissue genomes. *Am J Pathol* 163:37–45
8. Meissner A, Gnirke A, Bell GW et al (2005) Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res* 33: 5868–5877
9. Ammerpohl O, Martin-Subero JI, Richter J et al (2009) Hunting for the 5th base: techniques for analyzing DNA methylation. *Biochim Biophys Acta* 1790:847–862
10. Pogribny I, Yi P, James SJ (1999) A sensitive new method for rapid detection of abnormal methylation patterns in global DNA and within CpG islands. *Biochem Biophys Res Commun* 262:624–628

11. Fujiwara H, Ito M (2002) Nonisotopic cytosine extension assay: a highly sensitive method to evaluate CpG island methylation in the whole genome. *Anal Biochem* 307:386–389
12. Karimi M, Johansson S, Ekstrom TJ (2006) Using LUMA: a luminometric-based assay for global DNA-methylation. *Epigenetics* 1:45–48
13. Barres R, Osler ME, Yan J et al (2009) Non-CpG methylation of the PGC-1 alpha promoter through DNMT3B controls mitochondrial density. *Cell Metab* 10:189–198
14. El-Maarri O, Becker T, Junen J et al (2007) Gender specific differences in levels of DNA methylation at selected loci from human total blood: a tendency toward higher methylation levels in males. *Hum Genet* 122:505–514
15. El-Maarri O, Kareta MS, Mikeska T et al (2009) A systematic search for DNA methyltransferase polymorphisms reveals a rare DNMT3L variant associated with subtelomeric hypomethylation. *Hum Mol Genet* 18:1755–1768
16. Fish JE, Yan MS, Matouk CC et al (2010) Hypoxic repression of endothelial nitric-oxide synthase transcription is coupled with eviction of promoter histones. *J Biol Chem* 285:810–826
17. Perisic T, Zimmermann N, Kirmeier T et al (2010) Valproate and amitriptyline exert common and divergent influences on global and gene promoter-specific chromatin modifications in rat primary astrocytes. *Neuropsychopharmacology* 35:792–805
18. Pilsner JR, Lazarus AL, Nam DH et al (2010) Mercury-associated DNA hypomethylation in polar bear brains via the luminometric methylation assay: a sensitive method to study epigenetics in wildlife. *Mol Ecol* 19:307–314
19. Romermann D, Hasemeier B, Metzger K et al (2008) Global increase in DNA methylation in patients with myelodysplastic syndrome. *Leukemia* 22:1954–1956
20. Sadeghi RN, Zojaji H, Mohebbi SR et al (2009) Evaluation of global genome methylation in gastritis lesion and its correlation with clinicopathological findings. *Oncol Res* 17:549–558
21. Thompson RF, Atzmon G, Gheorghe C et al (2010) Tissue-specific dysregulation of DNA methylation in aging. *Aging Cell* 9:506–518
22. Thompson RF, Fazzari MJ, Greally JM (2010) Experimental approaches to the study of epigenomic dysregulation in ageing. *Exp Gerontol* 45:255–268
23. Zechner U, Nolte J, Wolf M et al (2009) Comparative methylation profiles and telomerase biology of mouse multipotent adult germline stem cells and embryonic stem cells. *Mol Hum Reprod* 15:345–353
24. Soriano-Tárraga C, Jiménez-Conde J, Giralt-Steinhaus E et al (2013) DNA isolation method is a source of global DNA methylation variability measured with LUMA. *Experimental analysis and a systematic review. PLoS One* 8:60750
25. Lind L, Penell J, Luttrupp K et al (2013) Global DNA hypermethylation is associated with high serum levels of persistent organic pollutants in an elderly population. *Environ Int* 59:456–461
26. Demetriou CA, Chen J, Polidoro S et al (2013) Methylome analysis and epigenetic changes associated with menarcheal age. *PLoS One* 8:79391
27. Head JA, Mittal K, Basu N et al (2014) Application of the luminometric methylation assay (LUMA) to ecological species; tissue quality requirements and a survey of DNA methylation levels in animals. *Mol Ecol Resour* 14:943–952
28. Ronaghi M (2001) Pyrosequencing sheds light on DNA sequencing. *Genome Res* 11:3–11
29. Kriaucionis S, Heintz N (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324:929–930
30. Booth MJ, Branco MR, Ficiz G et al (2012) Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* 336:934–937
31. Ichiyanagi K (2012) Inhibition of MspI cleavage activity by hydroxymethylation of the CpG site: a concern for DNA modification studies using restriction endonucleases. *Epigenetics* 7:131–136
32. Bjornsson HT, Sigurdsson MI, Fallin MD et al (2008) Intra-individual change over time in DNA methylation with familial clustering. *JAMA* 299:2877–2883
33. Fazzari MJ, Greally JM (2004) Epigenomics: beyond CpG islands. *Nat Rev Genet* 5:446–455

Free ebooks ==> www.Ebook777.com

Chapter 17

Limiting Dilution Bisulfite Pyrosequencing®: A Method for Methylation Analysis of Individual DNA Molecules in a Single or a Few Cells

Nady El Hajj, Juliane Kuhtz, and Thomas Haaf

Abstract

Bisulfite-based methods for DNA methylation analysis of small amounts of DNA from a limited number of cells are technologically challenging. Degradation of genomic DNA by bisulfite treatment, contamination with foreign DNA, and biases in the amplification of individual DNA molecules can generate results, which are not representative of the starting sample. Limiting dilution (LD) bisulfite Pyrosequencing® (BSP) is a relatively simple technique to circumvent these problems. The bisulfite-treated DNA of a single or a few cells is diluted to an extent, that only a single DNA target molecule is present in the reaction. Then each individual DNA molecule in the starting sample is separately amplified and analyzed by Pyrosequencing. This allows the detection of rare alleles that are easily masked when pools of DNA target molecules are analyzed. Amplicons containing a heterozygous single nucleotide polymorphism (SNP) allow one to delineate the parental origin of the recovered molecules in addition to their methylation status. The number of cells (DNA target molecules) in the starting sample determines the dilution level and the number of reactions that have to be performed. LD-BSP allows methylation analysis of small cell pools (i.e., 5–10 microdissected cells) and even individual cells. The primers and PCR conditions described here have been successfully employed to analyze the methylation status of up to eight target genes in individual 2–16 cell embryos, germinal vesicle (GV) oocytes, and haploid sperms.

Key words Allele-specific methylation, Bisulfite conversion, DNA methylation, Imprinted gene, Limiting dilution, Single cell analysis, Pyrosequencing®

1 Introduction

Bisulfite Pyrosequencing (BPS) is a bioluminometric real-time sequencing technique that allows quantitative DNA methylation measurements at single base resolution [1, 2]. Although it has several advantages, compared with other bisulfite based methods [3–5], methylation measurements can also be distorted by preferential amplification of either the methylated or the unmethylated alleles and stochastic amplification of only a few DNA molecules from the starting sample. Such an amplification bias

can lead to over/underrepresentation of specific DNA molecules in the sequencing reaction [6]. A general problem is epigenetic heterogeneity of the studied cell population. In many cases, the observed methylation differences between samples consisting of different cell types might be merely due to changes in cell composition rather than a specific treatment, disease status or other study question. Single-cell methylation analysis is often necessary for correct data interpretation.

Bisulfite sequencing remains the gold standard for DNA methylation analysis [7]. However, bisulfite treatment, which is necessary to distinguish between unmethylated cytosines that are converted into uracils and methylated cytosines which remain cytosines, heavily degrades DNA and makes it difficult to amplify. Bisulfite-converted DNA is largely depleted of cytosines. Reduced complexity complicates primer design [8, 9] and the high sequence divergence between methylated and unmethylated alleles after bisulfite treatment can promote preferential amplification of one allele type [6]. Methylation analysis of a few cells by bisulfite sequencing/Pyrosequencing is particularly challenging because the low amount of DNA in the starting sample augments the aforementioned problems [10]. When working with a limited number of cells, i.e., oocytes or early embryos, the results of methylation analyses have to be interpreted with caution. One strategy to avoid amplification bias is to pool at least several dozen cells, i.e., oocytes from different animals or different embryos [11–15]. Although many publications state that pools of 100 cells or so are large enough to get representative methylation measurements, in our experience there often is still a considerable amplification bias. Another strategy is to minimize DNA degradation by embedding single cells in agarose prior to bisulfite treatment [16, 17]. However, this protocol is labor intensive and so far has only been used successfully for methylation analysis of one target gene per cell. Whole genome amplification (WGA) of bisulfite-treated DNA [18] is currently not feasible in the picogram range. In our experience, generation of a representative WGA library requires at least one nanogram of bisulfite-treated DNA.

Recently, several techniques based on methylation-sensitive restriction enzymes were developed for single cell analysis. Since they do not rely on bisulfite treatment to distinguish between methylated and unmethylated cytosines, they are not influenced by DNA degradation. Restriction-based single-cell methylation assay (RSMA) applies the endonucleases *HpaII* and *Hin6I*, which cut the CCGG and GCGC recognition sequence, respectively, to single cells deposited on a microfluidic slide [19]. Another protocol uses *BstUI* digestion followed by multiplex real-time PCR for quantitative measurement of methylation values [20]. Restriction techniques offer an important alternative to bisulfite methods; however, they only assess the methylation status of a specific restriction site(s) with one or two CpGs per amplicon.

LD-BSP allows the simultaneous analysis of several neighboring CpG sites at single base resolution in individual DNA molecules [21] (*see Note 1*). This is crucial for differentiating between single CpG errors, which may represent stochastic events without functional implications or bisulfite conversion errors, and methylation abnormalities affecting the entire allele. Because it is usually the density of methylated CpGs in a cis-regulatory region that turns a gene “on” or “off” [22, 23], consistent allele-specific methylation errors can be considered as true epimutations, interfering with gene regulation. The principle of LD relies on diluting a bisulfite-converted DNA sample down to a level that either a single DNA target molecule or none is present in each reaction [24]. Since a given CpG on this target molecule is either methylated or not, the Pyrosequencing analysis has only to distinguish a binary state (Yes/No answer). Because quantitative measurement at a specific CpG site does not only depend on its methylation, but also on sequence context and other technical factors, for practical reasons, methylation values <20 % are indicative of an unmethylated site and >80 % for a methylated site [21, 25]. Values between 20 and 80 % can be due to the presence of several DNA target molecules in the reaction (due to inadequate dilution or DNA clumping) or to a technical artifact.

Single molecule analysis is crucial for studying methylation variation in a given cell population and for the identification of rare events, i.e., an epimutation in an individual cell within a larger population (needle-in-a-haystack). Importantly, SNPs with known parent-of-origin allow one to determine the methylation status of the paternal versus the maternal allele in a given cell [21]. The main advantage of LD-BSP compared to classical Pyrosequencing is avoidance of any amplification and cloning bias. In order to minimize the loss of target molecules, DNA isolation and bisulfite conversion are performed in the same tube. In our experience, the EZ-DNA methylation direct kit from Zymo Research (Irvine, CA, USA) is particularly suitable for this purpose. It includes a proteinase K DNA isolation step prior to bisulfite conversion. It is important to use reaction tubes and pipette tips which have a low DNA binding affinity. The dilution level depends on the amount of cells and DNA target molecules, respectively, in the starting sample (*see Note 2*). Following dilution most wells should contain either no or a single DNA target molecule, and only occasionally two or more molecules. Because of DNA degradation during bisulfite conversion, the number of amplifiable target molecules is usually much lower than the maximum number of DNA molecules in the starting sample. The number of recovered alleles can vary between different genes (amplicons) in a multiplex and between different cell substrates (*see Note 3*). On average, the recovery rate is 15–20 % of target molecules in the starting sample.

Prior to primer design, thorough in silico analysis should be performed to identify regions of interest containing strain-specific

SNPs (hybrid mice) or SNPs with a high rate of heterozygosity (human material). Primer design is the most challenging aspect of the technique. Primers must have a low potential for interacting with each other, a similar melting temperature, and a high binding efficiency/specificity for amplification of a single DNA target molecule. Multiplex assays require various optimization and validation steps. The time and costs involved in assay design may dramatically increase with the number of primer pairs in the multiplex. In principle, it is feasible to design multiplexes with 20 or more genes; however, we already experienced considerably lower amplification efficiencies for individual genes in assays with >8 primer pairs. The low complexity of bisulfite-converted DNA makes it increasingly difficult to add another primer pair to an existing multiplex without provoking primer interactions. The efforts needed to optimize and validate multiplex PCR for >8 targets are often bigger than working with two smaller multiplexes or singleplex reactions.

For LD-BSP it is of utmost importance to work in an environment which avoids any contamination with foreign DNA, especially during Subheadings 2.2–2.4 (*see* **Note 4**). We recommend doing all the pipetting in a sterile cabinet, which is decontaminated by ultraviolet light irradiation before starting work. In addition, we regularly decontaminate the bench by applying DNA-Exitus Plus (Applichem). Special precaution is necessary when working with a low number of human cells. This requires a UV bench which is not used for other purposes. To detect contamination artifacts, two to four water controls are added to the LD reactions for each analyzed sample.

2 Materials

2.1 Assay Design

1. PyroMark® assay design software (Qiagen) is used for PCR and sequencing primers.
2. MethPrimer (<http://www.urogene.org/cgi-bin/methprimer/methprimer.cgi>) is used for multiplex primer design [26].
3. Primer dimers and secondary structures are checked in silico using PriDimerCheck (http://biocompute.bmi.ac.cn/MPPrimer/primer_dimer.html).
4. BiSearch (<http://bisearch.enzim.hu>) is used for blasting bisulfite primers [9].

2.2 Cell Storage

1. The analyzed cells should be placed in 10 µL 1× PBS and stored at –80 °C until bisulfite conversion (*see* **Note 5**).

2.3 DNA Isolation and Bisulfite Conversion

1. Zymo EZ-DNA methylation direct kit with spin columns or, if a larger number of samples need to be analyzed, with 96-well plates (Zymo Research).

2. DNA LoBind tube (Eppendorf).
3. Low-retention sterile filter tips (Starlab) with repel polymer technology.

2.4 Multiplex PCR

1. Desalted primers.
2. AmpliTaq Gold polymerase (Applied Biosystems), FastStart Taq DNA polymerase (Roche Diagnostics), or PlatinumTaq DNA polymerase (Life Technologies).
3. 96-well plates and 8-well strip caps.
4. Low-retention sterile filter tips.
5. 96-well gradient thermal cycler with heated lid.

2.5 Nested PCR Amplification

1. Desalted primers.
2. HPLC purified 5' biotin labeled primers.
3. FastStart Taq DNA polymerase (Roche Diagnostics).
4. Sterile filter tips.
5. 8-Channel (0.5–10 µL) pipette.
6. 96-well gradient thermal cycler with heated lid.
7. 96-well plates and 8-well strip caps.

2.6 Gel Electrophoresis

1. Agarose.
2. Gel electrophoresis chamber.
3. DNA loading buffer.
4. 100 bp DNA ladder.

2.7 Pyrosequencing

1. PyroMark Q96 vacuum workstation.
2. Streptavidin Sepharose® high performance beads (GE Healthcare).
3. PyroMark binding buffer.
4. PyroMark annealing buffer.
5. 70 % Ethanol.
6. Denaturation buffer.
7. Washing buffer.
8. Pyrosequencing primers (standard desalting purification).
9. PyroMark Q96 HS plate.
10. Pyrosequencer.
11. Dispensation cartridges.
12. PyroMark Gold Q96 CDT reagents.

2.8 Analysis of the Pyrosequencing Run

1. PyroMark CpG software (Qiagen).

3 Methods

The different steps of LD-BSP are shown in Fig. 1 (single human oocyte analysis) and Fig. 2 (allele-specific methylation analysis of mouse hybrid two-cell embryos) (*see Note 6*). The primers for multiplex PCR analysis of seven human (*GTL2*, *LIT1*, *PEG3*, *SNRPN*, *NANOG*, *OCT4*, and *DNMT3Lo*) and four mouse (*H19*, *Igf2r*, *Snrpn*, and *Oct4*) genes, respectively, are presented in Tables 1 and 2. In addition, we have established multiplex assays for the analysis of human sperm (*GTL2*, *LIT1*, *PEG3*, *SNRPN*, *NANOG*, *OCT4*, *CATSPER1*, and *BOLL*), bovine oocytes (*H19*, *SNRPN*, and *PEG3*), and bovine embryos (*SLC2A1*, *PRDX1*, and *ZAR1*). Primer sequences and conditions for the latter assays are available upon request.

3.1 Assay Design

Primer design is a very crucial and challenging step that largely determines the success of LD-BSP. As mentioned earlier, bisulfite-treated DNA displays much lower complexity than genomic DNA. The PyroMark assay design software should be used for designing nested PCR and Pyrosequencing primers. MethPrimer (<http://www.urogene.org/cgi-bin/methprimer/methprimer.cgi>) is used for designing multiplex primers. Since outer primers have to be combined in the multiplex reaction, it is essential to check whether or not they can form primer dimers or secondary structures using appropriate software tools (http://biocompute.bmi.ac.cn/MPPrimer/primer_dimer.html).

3.1.1 Primer Design

1. The PyroMark assay design software is recommended for the design of nested (inner) primers. It takes into account secondary structures of primers and template DNA. Both the forward and reverse strand can serve as template (*see Note 7*); however, the same strand must be used for both outer and inner primers of a given gene in the multiplex. The primers should be 18–30 bp in length, have a melting temperature between 50 and 72 °C, and the amplification product size should not exceed 400 bp. These are default settings in the PyroMark assay design software. Several Pyrosequencing primers can be designed for the same amplicon to increase the number of analyzed CpGs. Usually, the read length of a sequencing primer is 60–70 bp. We recommend establishing the inner primers first.
2. MethPrimer is preferred for multiplex primer design. Although the default settings may be used, it is sometimes advisable to modify the parameters “maximum product size” and “primer non-CpGs.” Multiplex primers should have a similar melting temperature. To the extent possible, formation of primer dimers and secondary structures should be excluded by *in silico* analysis using PriDimerCheck. Because bisulfite DNA is highly degraded, amplicons of outer primers should not exceed 500 bp.

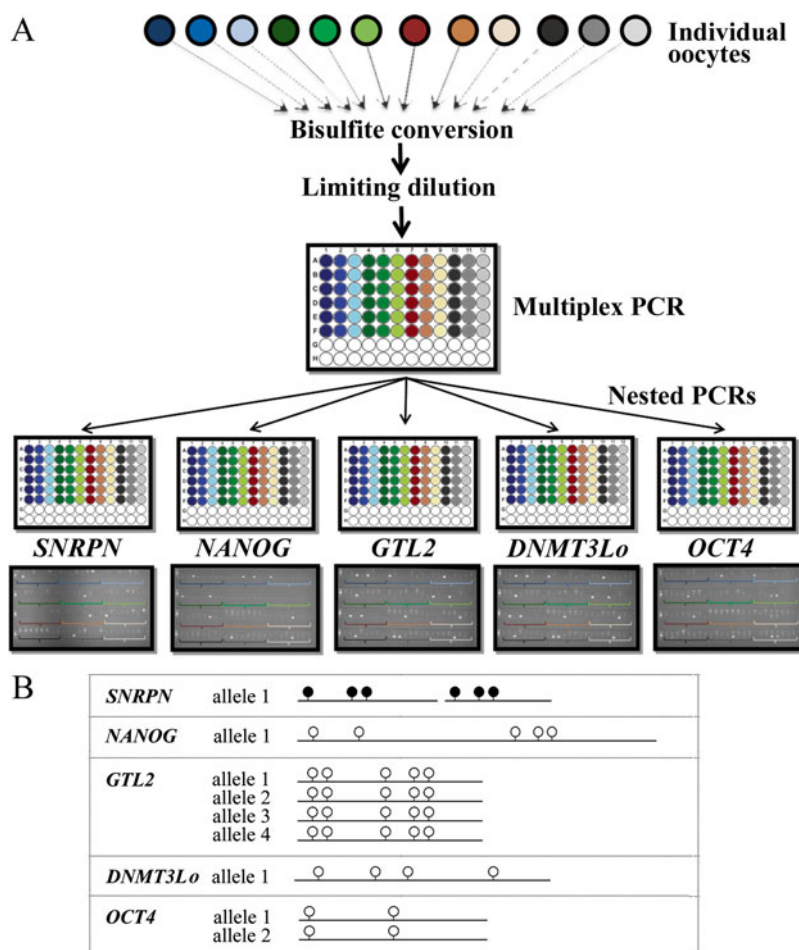


Fig. 1 (A) Flowchart of limiting dilution bisulfite Pyrosequencing of individual cells. In the depicted experiment, each of 12 individual human GV oocytes (indicated by different colors) is bisulfite-converted in a separate reaction. The bisulfite-treated DNA of a single oocyte (containing four DNA target molecules) is then diluted in six separate wells of a microtiter plate. Two wells per oocyte are used for water controls. Multiplex PCR of five target genes (*SNRPN*, *NANOG*, *GTL2*, *DNMT3Lo*, and *OCT4*) is performed in this plate. The first-round PCR product is then aliquoted in five different microtiter plates, one for second-round nested PCR of each target gene. The nested PCR products are resolved on gels to identify PCR reactions yielding a specific product from the single DNA molecule in the reaction. The color code marks all 30 nested PCR reactions (six per target gene) belonging to a specific oocyte. **(B)** The lollipop diagrams represent the methylation patterns (determined by bisulfite Pyrosequencing of nested PCR products) of all alleles recovered from one individual oocyte. *Open circles* represent unmethylated and *filled circles* methylated CpG sites. Since each PCR product was amplified in a separate reaction of diluted DNA, all four *GTL2* alleles in the starting sample were recovered. The nine recovered alleles (of five genes) show normal oocyte methylation patterns. *GTL2* is a paternally methylated and *SNRPN* a maternally methylated imprinted gene. The pluripotency genes *OCT4* and *NANOG* as well as the oocyte-specific DNA methyltransferase *DNMT3Lo* are unmethylated in oocytes but methylated in somatic cells

As a rule of thumb, the smaller the amplicon size the higher the amplification efficiency. The most important point is to have primers with high binding affinity, which can generate a specific product from the single DNA molecule in the reaction.

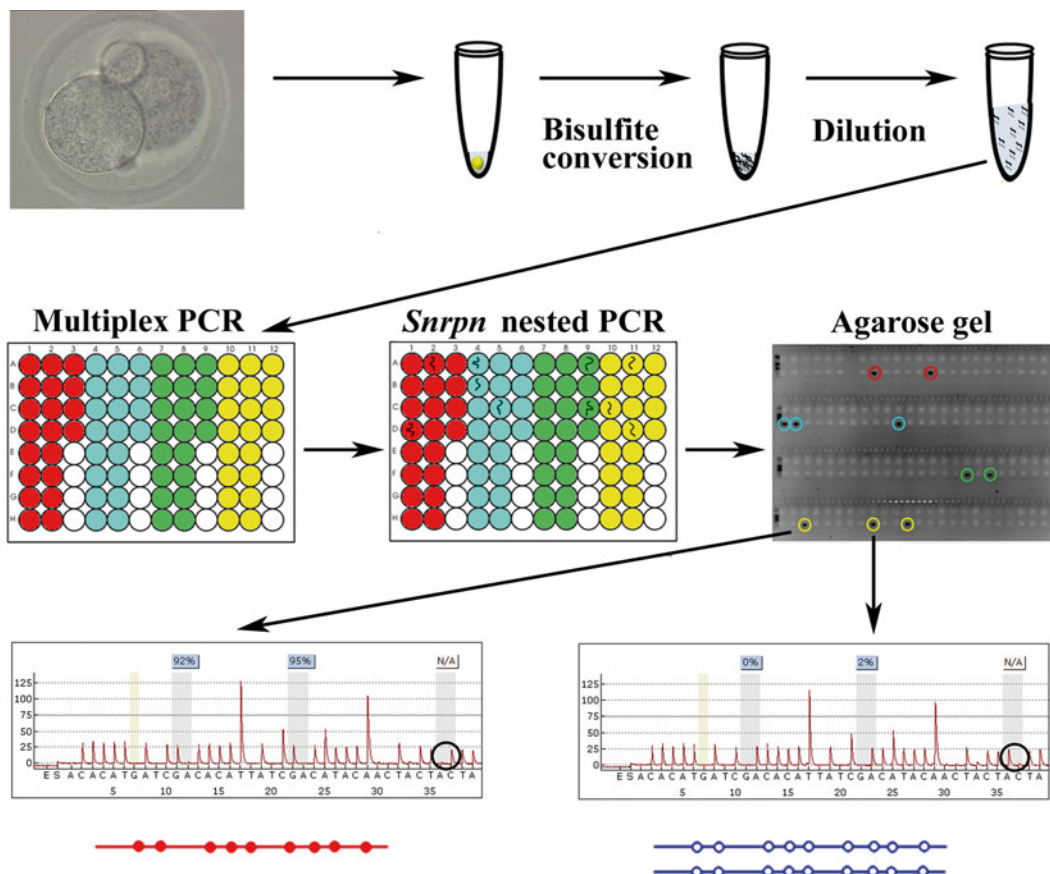


Fig. 2 Allele-specific LD bisulfite Pyrosequencing. An individual mouse hybrid two-cell embryo without contaminating somatic cells is transferred to a reaction tube for bisulfite conversion and dilution. Two-cell embryo and retained second polar body contain a maximum of nine DNA target molecules. The bisulfite-converted DNA of an individual embryo (indicated by *yellow color*) is distributed in 20 wells of a microtiter plate. Four water controls are added for each of the four analyzed embryos (indicated by different colors). Multiplex PCR for three (paternally methylated *H19* and maternally methylated *Snrpn* and *Igf2h*) imprinted genes and the pluripotency gene *Oct4* and second round nested PCRs are performed in microtiter plates and the resulting PCR products are checked on agarose gels for the presence of specific amplification products. For the sake of simplicity, only one of the four second-round nested PCR plates and gels is shown. The three *Snrpn* products recovered from the *yellow* embryo were then analyzed by Pyrosequencing. In addition to two CpG sites, the amplicon contains a strain-specific SNP (*encircled*) which allows one to distinguish between maternal (*Mus musculus*) and paternal (*M. musculus castaneus*) alleles. The one recovered maternal allele was fully methylated (indicated by filled *red dots*) and the two paternal alleles (indicated by open *blue circles*) were demethylated, consistent with normal *Snrpn* imprints in this embryo. Two different sequencing primers were used for the same amplicon, which together assess nine neighboring CpG sites (*see Table 2*)

3. The primers should not include CpGs in their binding site since after bisulfite conversion an unmethylated cytosine is transformed to uracil whereas the methylated cytosine will remain unchanged. In case this cannot be avoided, then the CpG should be positioned at the 5' end of the primer sequence and a wobble base included instead of the cytosine.

Table 1
Primers for a multiplex with human imprinted (*GTL2*, *LIT1*, *PEG3*, and *SNRPN*) and non-imprinted (*NANOG*, *OCT4*, and *DNMT3L*) genes

Gene		Primer sequence	Annealing temp. (°C)	Cycle number	Polymerase	Number of CpGs
<i>GTL2</i>	Outer forward	GAATTATAGGGAATGATGGTGTA	58	35×	FastStart	5
	Outer reverse	CCAAAATACTAACTACTCTTAAACA				
	Nested forward	AGGGTAGGAAGTTTAGTAGGTTA	60	32×	FastStart	
	Nested reverse	ACTACTCCTTAAACAAAAACACATAAT ^a				
	Pyrosequencing	GTAGTAAATTAAAGTGTATTAGAGA				
<i>LIT1</i>	Outer forward	GGGAGAATAGTGTTGAGGAGTTT	58	35×	FastStart	6
	Outer reverse	CCCTACTACCCCCCAATCAACAAATAA				
	Nested forward	GGGGAGGATTAAAGTTGAGAGGTA	57	35×	FastStart	
	Nested reverse	CCCTACTACCCCCCAATCAACAAATAA ^a				
	Pyrosequencing	GGAGTTTTTTGGAGG				
<i>PEG3</i>	Outer forward	GGTTGTTGATTGGTTAGTATAG	58	35×	FastStart	3
	Outer reverse	CACTCACCTCACCTCAATAC				
	Nested forward	GGTGTAGAAGTTTGGGTAGTTG	60	32×	FastStart	
	Nested reverse	CTCACCTCACCTCAATACTAC ^a				
	Pyrosequencing	TGTTTAITTTTGGGTGGT				
<i>SNRPN</i>	Outer forward	GGGTTTTAGGGGTTTAGTAGT	58	35×	FastStart	3
	Outer reverse	ACTCCAAATCCTAAAAACTTAAATATC				
	Nested forward	AGGGAGTTGGGATTTTGTATT ^a	60	32×	FastStart	
	Nested reverse	CCCAAACTATCTCTTAAAAAAAC				
	Pyrosequencing 1	ACACAACCTAACCTTACCC				
	Pyrosequencing 2	CCAACCTACCTCTAC				

(continued)

Table 1
(continued)

Gene		Primer sequence	Annealing temp. (°C)	Cycle number	Polymerase	Number of CpGs
<i>NANOG</i>	Outer forward	TAGAGTAATTTAGATTAGGTGGGGAAT	58	35×	FastStart	
	Outer reverse	TAACCAAACTAATTTCAAACTCCTAACTT				
	Nested forward	GGAATATGGTTAAATAGGAATGGGATAA	60	32×	FastStart	
	Nested reverse	ACCAAACTAATTTCAAACTCCTAACTTCA ^a				2
	Pyrosequencing 1	ATTTTGTAAATTTAGTAAATTTGGG				3
	Pyrosequencing 2	TTTTTAAAAATTAAGAAAAAGGT				
<i>OCT4</i>	Outer forward	AGAAGGATTGTTTGTGGTTTAGTAGA	58	35×	FastStart	
	Outer reverse	AAC TACTCAACCCCTCTCT				
	Nested forward	AAGTTTTTGTGGGGGATTGTAT	60	32×	FastStart	
	Nested reverse	CCACCCACTAACCTTAAACCTCTA ^a				
	Pyrosequencing	TGAGGTTTTGGAGGG				2
<i>DNMT3Lo</i>	Outer forward	GGAAGTGAGAGTTTTTTTGAGT	58	35×	FastStart	
	Outer reverse	ATCAAAACCAAAACCTCCC				
	Nested forward	TGTGTTTGGTGGGTGAGT	60	32×	FastStart	
	Nested reverse	ATCAAAACCAAAACCTCCC ^a				
	Pyrosequencing	GTGGGTGAGTTAGGTTA				4

^aBiotinylated primer

Table 2
Primers for mouse *H19*, *Igf2r*, *Snrpn*, and *Oct4* multiplex assay

Gene	Primer sequence	Annealing temp. (°C)	Cycle number	Polymerase	Number of CpGs
<i>H19</i>	Outer forward	52	35×	Amplitaq Gold	
	Outer reverse				
	Nested forward	57	35×	FastStart	
	Nested reverse				
	Pyrosequencing 1				4
	Pyrosequencing 2				1
<i>Igf2r</i>	Outer forward	52	35×	Amplitaq Gold	
	Outer reverse				
	Nested forward	57	37×	FastStart	
	Nested reverse				
	Pyrosequencing 1				4
	Pyrosequencing 2				2
<i>Snrpn</i>	Outer forward	52	35×	Amplitaq Gold	
	Outer reverse				
	Nested forward	55	29×	FastStart	
	Nested reverse				
	Pyrosequencing 1				2
	Pyrosequencing 2				7
<i>Oct4</i>	Outer forward	52	35×	Amplitaq Gold	
	Outer reverse				
	Nested forward	57	38×	FastStart	
	Nested reverse				

^aBiotinylated primer

4. All primers should be tested in silico for unspecific amplification. BiSearch is helpful to search bisulfite-adapted genomes for possible amplification products on the sense and antisense strand.
5. The primer binding sites should be checked for the presence of SNPs which can lead to preferential amplification of an allele or failed amplification, especially when the SNP is towards the 3' end of the primer sequence.
6. When studying parent-of-origin effects (i.e., in mouse hybrid embryos), SNPs should be identified prior to assay design using an in silico analysis tool (*see Note 8*).

3.1.2 Primer Testing

1. For determining the optimal annealing temperature, each (outer and inner) primer pair has to be tested separately using FastStart Taq DNA polymerase, 10 ng of bisulfite converted DNA, and 40 cycles. A specific product should be observed for all pairs. For primer testing, a low reaction volume of 12.5 μ L reduces costs for reagents.
2. Then all primer pairs are combined and multiplex PCR is performed using AmpliTaq Gold polymerase, FastStart Taq DNA polymerase, PlatinumTaq DNA polymerase, or the Qiagen Multiplex PCR kit. Different PCR systems may lead to variable results. The multiplex reaction should be tested with a gradient for annealing temperature in order to define conditions which yield specific products without primer interactions. Usually, a gradient between 52 and 60 °C is sufficient; however, in some cases the optimum annealing temperature may be lower. It is essential that primers do not form secondary structures or primer-dimers which decrease their amplification efficiency.
3. When the multiplex works well with standard amounts of genomic DNA, then the efficiency is tested using lower DNA concentrations, i.e., a 1:10 and 1:100 dilution of 10 ng bisulfite-converted DNA.
4. There are several possibilities when no PCR product(s) are observed. In most cases the problems are associated with outer primers in the multiplex PCR reaction. A number of adjustments can be tried out to improve PCR efficiency, including:
 - Different $MgCl_2$ concentrations.
 - Testing different primer combinations via semi-nested PCR where the outer primer is replaced by the respective forward or reverse inner primer.
 - Different Taq concentrations in the PCR reaction.
 - Different primer concentrations.
 - Adding 5–10 % dimethyl sulfoxide (DMSO) to the PCR reaction.

- Adding up to 0.8 µg/µL bovine serum albumin (BSA) to the PCR reaction.

In case the above mentioned adjustments do not improve PCR efficiency, new primers have to be designed.

5. After determining the optimal PCR conditions, a test run is performed with low amounts (0.2–1 ng) of artificially methylated and unmethylated genomic DNA. Multiplex PCR and nested PCRs are performed without dilution. DNA standards can be purchased from different providers, e.g., EpiTect Control DNA (Qiagen) or prepared in the lab. Unmethylated DNA is created by WGA with Phi29 polymerase and fully methylated DNA by treatment of DNA with the CpG methyltransferase SssI.

3.2 DNA Isolation and Bisulfite Conversion

The EZ DNA methylation direct Kit (Zymo Research) is used for DNA isolation and bisulfite conversion. Cells stored in 10 µL 1× PBS at –80 °C are thawed and processed further, as detailed below:

1. Shortly centrifuge the tubes containing the thawed cells (*see Note 9*) and add 10 µL 2× M-digestion buffer and 1 µL proteinase K.
2. Flick tube several times to mix its content. Mixing by pipetting might lead to cell loss (*see Note 10*).
3. Incubate samples for 20 min at 50 °C for digestion.
4. Add 130 µL CT-conversion reagent to the tube, flick (mix) and then shortly spin tubes.
5. Place the reaction in a thermal cycler and incubate at 98 °C for 8 min and then at 64 °C for 210 min.
6. If necessary, samples can be stored at 4 °C for up to 20 h.
7. Load 600 µL M-binding buffer on the Zymo-Spin IC column and add the whole sample volume from the previous step with a low-retention sterile filter tip.
8. Close spin column and invert several times to mix the content.
9. Centrifuge for 30 s at full speed and then discard the flow-through.
10. Add 100 µL M-wash buffer to the Zymo-Spin IC column and centrifuge for 30 s at full speed.
11. Add 200 µL of M-desulfonation buffer to the spin column and leave for 15 min at room temperature. Centrifuge at full speed for 30 s and then discard flow-through.
12. Add 200 µL M-wash buffer and centrifuge for 30 s at full speed. Repeat once.
13. Transfer the column to a new 1.5 mL tube and add 10 µL of M-elution buffer to the center of the membrane. Centrifuge at full speed for 30 s.

14. The bisulfite treated DNA can be stored at -20°C for up to 1 week; however, the best results are achieved without further storage (*see* **Note 11**).

3.3 Multiplex PCR

It is generally worth to test different PCR systems to find the optimum polymerase and conditions for a given primer library. Generating a product from a single DNA target molecule requires a sufficient number of cycles. On average we use 35 cycles in the multiplex PCR and 29–38 cycles in the subsequent nested singleplex PCRs. Dilution of the starting sample and, thus, the number of reactions depends on the maximum number of DNA targets in the sample (*see* **Note 2**). Multiplex PCR is performed in 25 μL reactions; however, a 12.5 μL reaction system can be used to reduce costs.

1. Dilute 10 μL of bisulfite-treated DNA appropriately, i.e., 1:6 for a single oocyte or 1:20 for a two-cell embryo. Each multiplex reaction should contain 10 μL of the diluted DNA. Thus, if a 1:20 dilution is required, 190 μL water are added to 10 μL of bisulfite treated DNA and 10 μL aliquots are pipetted into 20 wells of a microtiter plate using a sterile low-retention tip.
2. Two or better four water controls per sample are added to detect DNA contamination during amplification. For this, 10 μL ultrapure water each are pipetted into wells of the microtiter plate using a new low retention tip.
3. The master mix for the multiplex PCR is then prepared using the polymerase system of choice and the multiplex PCRs are amplified using the preset cycling conditions.

3.4 Nested PCR

Following multiplex PCR, nested singleplex PCRs are performed for each gene using 1 μL of the multiplex PCR product as template. The standard protocol employs FastStart Taq polymerase in a 25 μL volume.

1. 1 μL multiplex PCR product each is transferred from the 96-well multiplex PCR plate into a new plate. Using a 8-channel pipette for plate to plate transfer decreases hands-on time.
2. The pipetting scheme includes 2.5 μL 10 $\times$ buffer, 0.5 μL dNTP mix, 1.25 μL of the forward and 1.25 μL reverse primer (10 pmol/ μL), 0.2 μL Fast Start Taq polymerase, and 18.3 μL PCR grade water per reaction. A master mix is prepared and 24 μL aliquots are distributed into the wells of a microtiter plate using a multi-dispenser.
3. The cycling conditions for the second-round nested PCR may vary from gene to gene and are determined for each gene (amplicon) separately.

4. The resulting nested PCR products are resolved on 1–2 % agarose gels to identify reactions (wells) containing a specific amplification product, i.e., a band of the right size (*see* **Note 12**). Loading buffer is added to 8-strip format PCR tubes and then multichannel pipettes are used for loading the needed volume into a 96-well plate. An 8-channel pipette is used for pipetting 4 μL of each PCR product to the loading buffer. The negative control wells should be carefully checked for contamination (*see* **Note 13**).

3.5 Pyrosequencing

The identified specific amplification products are analyzed by Pyrosequencing, using a Pyromark MD or comparable system.

1. A master mix is prepared for the required number of Pyrosequencing reactions. The reaction volumes for a single sample include 40 μL binding buffer, 28 μL water, and 2 μL streptavidin Sepharose high performance beads.
2. Vortex the master mix and add 70 μL each into the required number of wells in a microtiter plate using a dispenser pipette.
3. Transfer 10 μL of nested PCR product (gel verified) into the wells containing the master mix, seal the plate with a cover tape, and vortex for 5 min.
4. In the meantime, add 11.5 μL of annealing buffer and 0.5 μL of sequencing primer to the wells of a PyroMark Q96 HS plate.
5. Clean the filter tips of the Vacuum Prep Tool (VPT) in a trough with ultrapure water and then immerse the VPT into the plate containing Sepharose beads with immobilized PCR product and leave for a few seconds.
6. Sequentially submerge the head into the following troughs: 5 s in 70 % ethanol, 5 s in denaturation buffer, and 5–10 s in washing buffer.
7. Place the vacuum head on top of the sequencing plate, turn off the vacuum, and lower the VPT into the PyroMark Q96 HS plate. Shake gently with concentric motion to release the Sepharose beads bound to single-stranded DNA.
8. Place the sequencing plate on a heating block at 80 °C for 2 min.
9. Fill the dispensation cartridge with the indicated volume of enzyme, substrate, and nucleotides.
10. Allow the PyroMark Q96 HS plate to cool down for 5 min and then place it in the Pyromark MD instrument.
11. Start the pyrosequencing run.

3.6 Analysis of Pyrosequencing Results

Data analysis is performed using PyroMark CpG software. The methylation levels are measured at single CpG resolution and the nucleotide at the variable SNP position is identified. LD analyzes

single DNA target molecules in the amplification reaction, and thus, the methylation measured at individual CpGs should be either 0 % or 100 %. However, nested PCR usually increases background noise. In our experience, it is safe to classify values <20 % as unmethylated and values >80 % as fully methylated CpGs. Alleles with intermediate methylation values from 20 to 80 % should be excluded from further analysis. The Pyrograms of the water-template wells should be carefully controlled for contamination (*see* **Note 13**).

4 Notes

1. Apart from Pyrosequencing, the LD products can also be analyzed by direct dideoxy sequencing. This is more labor and time intensive, but allows one to look at a larger number of CpGs. Direct sequencing primers are M13 tagged and include stuffer sequences comprised of 50 % cytosine in the forward primer and 50 % guanine in the reverse primer. Stuffer sequences are required to improve the sequence quality of the less complex bisulfite DNA [27, 28].
2. The dilution level is critical for having a single DNA target molecule per well. Employing a Poisson distribution can be helpful; however, in most cases the optimum dilution is determined empirically. As a rule of thumb, if most wells contain at most one DNA target molecule, less than 50 % of the wells of second round nested PCRs can be expected to generate a product. We use a 1:6 dilution for an individual oocyte (including first polar body), a 1:20 dilution for a pool of 10 oocytes or a single two-cell embryo, and a 1:90 dilution for a 16-cell embryo. Single sperm DNA is not diluted at all.
3. The allele recovery rate in a multiplex analysis of seven genes in human GV oocytes (containing four DNA target molecules each) was 28 % for *GTL2*, 3.3 % for *LIT1*, 11.7 % for *PEG3*, 6.8 % for *SNRPN*, 20.5 % for *NANOG*, 11.7 % for *OCT4*, and 21.8 % for *DNMT3Lo*. Using another multiplex with four genes for the analysis of mouse two-cell embryos, the average recovery rate was 25 % for *H19*, 11 % for *Igf2r*, 19 % for *Snrpn*, and 21 % for *Oct4*, assuming that half of the blastomeres are in G1 phase and half have already replicated [21].
4. It is very important to collect the studied cells in a sterile environment and to avoid any DNA contamination during all further steps. Oocytes or early embryos are easily contaminated with somatic DNA from damaged surrounding cumulus cells. Somatic DNA methylation patterns due to unintended contamination would be falsely interpreted as allele methylation errors in the oocyte. When studying oocytes, it is recommended to include 1–2 marker genes for somatic DNA in the

multiplex, i.e., the pluripotency markers *OCT4* and *NANOG* are unmethylated in oocytes and methylated in somatic cells.

5. Cells can be stored in 1× PBS at −80 °C. We have worked with cells that were frozen for several months without decreasing the LD efficiency. DNA LoBind tubes should be used for freezing and storage.
6. Imprinted genes are essential for the regulation of fetal and placental growth, somatic differentiation as well as neurological and behavioral functions. They are frequently used as a model to assess the epigenetic effects of assisted reproductive technologies during germ-cell and embryo development. Parental allele-specific methylation of cis-regulatory elements (differentially methylated imprinting control regions) during gametogenesis controls the preferential expression of imprinted genes from one of the two parental alleles after fertilization [29, 30]. *LIT1*, *PEG3*, *SNRPN/Snrpn*, and *Igf2r* are methylated in the maternal germ line, whereas *H19* and *GTL2* are methylated in the paternal germ line.
7. The sequence to analyze should not contain homopolymer stretches, which may interfere with quantification accuracy of Pyrosequencing. Homopolymer stretches can be avoided by designing primers for the opposite DNA strand. After bisulfite conversion, sense and antisense strand are no longer complementary.
8. When amplicons contain a heterozygous SNP for parent-of-origin determination, it must be checked whether the site is still polymorphic after bisulfite treatment. For example, a C/T SNP gives a thymine on both alleles after conversion. In this case, the opposite strand with an A/G SNP can be used, which remains after conversion.
9. Before applying the EZ DNA methylation direct kit for DNA isolation and bisulfite conversion, a short centrifugation step helps to collect all cells and liquid at the bottom of the reaction tube.
10. Harsh vortexing during DNA isolation and bisulfite conversion inevitably leads to loss of DNA target molecules. To mix reagents, mild vortexing or better tube flicking can be used.
11. If possible, methylation analysis should be performed shortly (within 1 week) after bisulfite conversion. The multiplex PCR products can be stored at −20 °C for several months without reducing the efficiency of second-round nested PCRs.
12. The percentage of wells yielding a specific PCR product is important to judge the quality of the experiment. Following appropriate dilution of the starting sample and without DNA contamination, the majority of wells should not generate a product.

13. For each studied sample, several water controls (without template DNA) should be processed in parallel, to detect DNA contamination. When amplification products are seen in one or a few control wells, the nested PCR should be repeated. If the problem persists, most likely the starting DNA sample or the multiplex product has already been contaminated, and this sample must be excluded from analysis.

References

1. Ronaghi M (2001) Pyrosequencing sheds light on DNA sequencing. *Genome Res* 11:3–11
2. Ronaghi M, Uhlén M, Nyrén P (1998) A sequencing method based on real-time pyrophosphate. *Science* 281:363–365
3. Colella S, Shen L, Baggerly KA et al (2003) Sensitive and quantitative universal pyrosequencing methylation analysis of CpG sites. *Biotechniques* 35:146–150
4. Tost J, Dunker J, Gut IG (2003) Analysis and quantification of multiple methylation variable positions in CpG islands by pyrosequencing. *Biotechniques* 35:152–156
5. Uhlmann K, Brinckmann A, Toliat M et al (2002) Evaluation of a potential epigenetic biomarker by quantitative methyl-single nucleotide polymorphism analysis. *Electrophoresis* 23:4072–4079
6. Warnecke PM, Stirzaker C, Melki JR et al (1997) Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. *Nucleic Acids Res* 25:4422–4426
7. Frommer M, McDonald LE, Millar DS et al (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad U S A* 89:1827–1831
8. Arányi T, Váradi A, Simon I et al (2006) The BiSearch web server. *BMC Bioinformatics* 7:431
9. Tusnády GE, Simon I, Váradi A et al (2005) BiSearch: primer-design and search tool for PCR on bisulfite-treated genomes. *Nucleic Acids Res* 33:e9
10. Lucifero D, Mertineit C, Clarke HJ et al (2002) Methylation dynamics of imprinted genes in mouse germ cells. *Genomics* 79:530–538
11. Hiura H, Obata Y, Komiyama J et al (2006) Oocyte growth-dependent progression of maternal imprinting in mice. *Genes Cells* 11: 353–361
12. Anckaert E, Romero S, Adriaenssens T et al (2010) Effects of low methyl donor levels in culture medium during mouse follicle culture on oocyte imprinting establishment. *Biol Reprod* 83:377–386
13. Tomizawa S, Kobayashi H, Watanabe T et al (2011) Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. *Development* 138:811–820
14. Fauque P, Jouannet P, Lesaffre C et al (2007) Assisted reproductive technology affects developmental kinetics, H19 imprinting control region methylation and H19 gene expression in individual mouse embryos. *BMC Dev Biol* 7:116
15. Market-Velker BA, Zhang L, Magri LS et al (2010) Dual effects of superovulation: loss of maternal and paternal imprinted methylation in a dose-dependent manner. *Hum Mol Genet* 19:36–51
16. Hajkova P, El-Maarri O, Engemann S et al (2002) DNA-methylation analysis by the bisulfite-assisted genomic sequencing method. *Methods Mol Biol* 200:143–154
17. Geuns E, de Rycke M, van Steirteghem A et al (2003) Methylation imprints of the imprint control region of the SNRPN-gene in human gametes and preimplantation embryos. *Hum Mol Genet* 12:2873–2879
18. Mill J, Petronis A (2009) Profiling DNA methylation from small amounts of genomic DNA starting material: efficient sodium bisulfite conversion and subsequent whole-genome amplification. *Methods Mol Biol* 507:371–381
19. Kantlehner M, Kirchner R, Hartmann P et al (2011) A high-throughput DNA methylation analysis of a single cell. *Nucleic Acids Res* 3:e44
20. Lorthongpanich C, Cheow LF, Balu S et al (2013) Single-cell DNA-methylation analysis reveals epigenetic chimerism in preimplantation embryos. *Science* 341:1110–1112
21. El Hajj N, Trapphoff T, Linke M et al (2011) Limiting dilution bisulfite (pyro)sequencing reveals parent-specific methylation patterns in

- single early mouse embryos and bovine oocytes. *Epigenetics* 6:1176–1188
22. Sontag LB, Lorincz MC, Luebeck GE (2006) Dynamics, stability and inheritance of somatic DNA methylation imprints. *J Theor Biol* 242: 890–899
 23. Weber M, Hellmann I, Stadler MB et al (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet* 39:457–466
 24. Daser A, Thangavelu M, Pannell R et al (2006) Interrogation of genomes by molecular copy-number counting (MCC). *Nat Methods* 3: 447–453
 25. Trapphoff T, El Hajj N, Zechner U et al (2010) DNA integrity, growth pattern, spindle formation, chromosomal constitution and imprinting patterns of mouse oocytes from vitrified pre-antral follicles. *Hum Reprod* 25:3025–3042
 26. Li LC, Dahiya R (2002) MethPrimer: designing primers for methylation PCRs. *Bioinformatics* 18:1427–1431
 27. Diederich M, Hansmann T, Heinzmann J et al (2012) DNA methylation and mRNA expression profiles in bovine oocytes derived from prepubertal and adult donors. *Reproduction* 144:319–330
 28. Heinzmann J, Hansmann T, Herrmann D et al (2011) Epigenetic profile of developmentally important genes in bovine oocytes. *Mol Reprod Dev* 78:188–201
 29. Reik W, Constância M, Fowden A et al (2003) Regulation of supply and demand for maternal nutrients in mammals by imprinted genes. *J Physiol* 547:35–44
 30. Kelsey G (2007) Genomic imprinting – roles and regulation in development. *Endocr Dev* 12: 99–112

Free ebooks ==> www.Ebook777.com

Chapter 18

Detection of Loss of Imprinting by Pyrosequencing®

Silvia Tabano, Eleonora Bonaparte, and Monica Miozzo

Abstract

Genomic imprinting is an epigenetically regulated process determining allele-specific expression in a parent-of-origin dependent manner. Altered expression of imprinted genes characterizes numerous congenital diseases including Beckwith–Wiedemann, Silver–Russell, Angelman, and Prader–Willi syndromes as well as acquired disorders such as cancer. The detection of imprinting alterations has important translational implications in clinics and the application of the Pyrosequencing® technology offers the possibility to identify accurately also subtle modifications in allele-specific expression and in DNA methylation levels.

Here, we describe two methods to investigate genomic imprinting defects (loss of imprinting, LOI) using Pyrosequencing: (1) Allele-specific expression analysis based on single nucleotide polymorphism (SNP), and (2) quantification of DNA methylation.

The protocol for the quantification of the allele-specific expression is carried out by analyzing an informative SNP located within the transcribed portion of an imprinted gene. The method includes the cDNA amplification of the region containing the SNP and the Pyrosequencing-based analysis for the quantitative allelic discrimination comparing the ratio of the two alleles.

The second protocol allows the accurate quantification of the DNA methylation levels at the Imprinting Control Regions (ICRs). Imprinted genes are clustered in chromosomal regions and their expression is mainly regulated by DNA methylation at CpG sites located within the ICRs. After bisulfite modification of the genomic DNA, the region of interest is amplified by PCR and analyzed by Pyrosequencing. The methylation value at each CpG site is calculated by the CpG software, which determines the ratio of the incorporation of “C” and “T” and converts the value in methylation percentage.

Key words Genomic imprinting, Epigenetics, DNA methylation, Pyrosequencing®, Allele-specific expression, Loss of imprinting, ICR, Prader–Willi syndrome, Angelman syndrome, Beckwith–Wiedemann syndrome, Silver–Russell syndrome

1 Introduction

In human, aberrant genomic imprinting underlies a number of congenital disorders (i.e., genomic imprinting syndromes) and is common in cancer.

Loss of imprinting (LOI) is defined as a parent-of-origin specific epigenetic modification that is disrupted and includes gain or loss of methylation or loss of the normal pattern of parent-of-origin specific gene expression pattern. Therefore, LOI comprises

activation of the normally silent copy (e.g., *Insulin-like Growth Factor 2*, *IGF2*), or silencing of the normally active copy (e.g., *H19*) of an imprinted gene [1].

LOI was first discovered in Wilm's tumor [2], which shows relaxation of the normal silencing of the maternal allele of *IGF2* and, consequently, biallelic expression of the gene. Subsequently, LOI of *IGF2* was commonly found in tumors, giving evidence that *IGF2* is a major factor contributing to oncogenesis [3].

Congenital disorders related to genomic imprinting defects (e.g., Beckwith–Wiedemann, Silver–Russell, Prader–Willi, and Angelman syndromes) are associated with heterogeneous epigenetic and genetic alterations, most of them leading to changes of the expected methylation pattern [4]. The detection of epigenetic defects by reliable approaches such as Pyrosequencing® is very useful in research studies as well as for molecular diagnosis of diseases related to genomic imprinting alterations.

Pyrosequencing is based on the incorporation of a nucleotide into the template strand that leads to the release of pyrophosphate, which is quantified by a luciferase reaction. The light signal is proportional to the amount of pyrophosphate released. Consequently, SNPs as well as the sequence modified by the chemically induced C/T conversion can be detected and precisely quantified.

1.1 Protocol for the Detection of LOI by SNP Analysis

LOI assay can be performed by Pyrosequencing through the quantification of the two expressed alleles by analyzing at least an informative single nucleotide polymorphism (SNP) located within the transcribed portion of an imprinted gene, and therefore present in the mRNA. By definition, the expression of imprinted genes is monoallelic, and therefore, in the presence of a normal imprinting, it is expected to find an allele with 100 % of expression and the other one almost completely silenced (0 %).

Starting from cDNA amplification of the region containing the informative (heterozygous) SNP, Pyrosequencing enables quantitative allele discrimination and allows distinguishing the two allelic expression patterns by comparing the ratio of the two alternative alleles (Fig. 1).

In order to exactly quantify the contribution of each allele to gene expression, it is possible to normalize the raw expression data by interpolation with a calibration curve generated with the help of standard samples. “Standard samples” are obtained by mixing two DNA samples at different proportions that are each homozygous for one of the two genotypes under investigation.

The feasibility of this technique depends on: (1) the presence of informative SNPs in the coding sequence of the gene under study; (2) Availability of RNA: the analysis is based on cDNA sequencing, and therefore, RNA must be available.

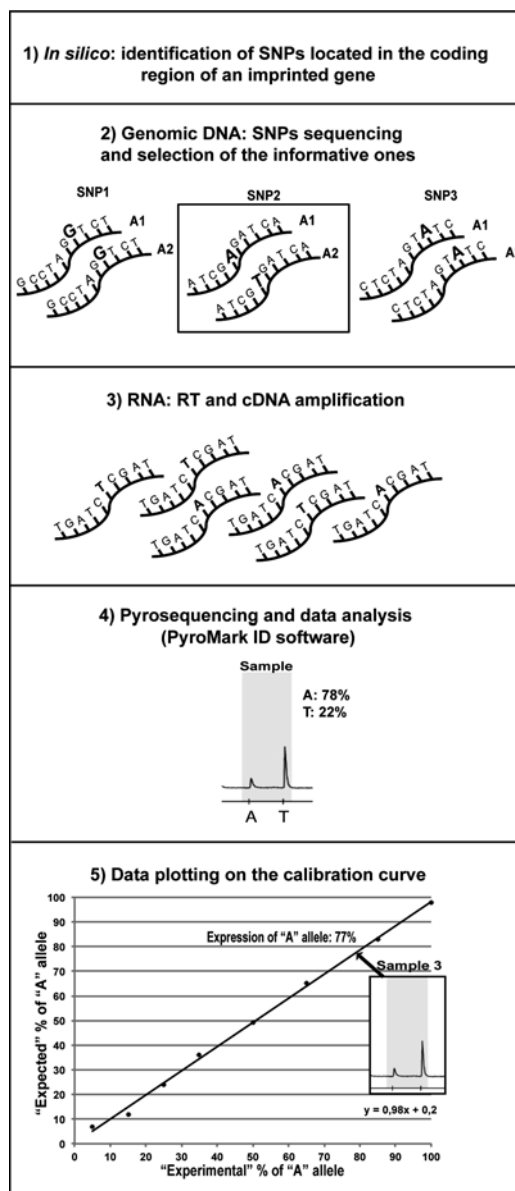


Fig. 1 Schematic representation of the protocol to identify LOI by analyzing coding SNPs. (1) SNPs are selected in silico, based on % of heterozygosity and position on the coding sequence of an imprinted gene. (2) SNPs-containing regions of genomic DNA are amplified and sequenced to identify the informative SNPs. (3) RT-PCR and (4) Pyrosequencing experiments are performed on heterozygous SNPs. (5) A calibration curve is generated by comparing the “expected” (based on sample composition) with the “experimental” (based on allele quantification by the PyroMark ID system) allelic percentages of standard samples (e.g., two samples homozygous for the ancestral and the mutant allele, respectively). By interpolating the “experimental” expression values of an unknown sample with the calibration curve, it is possible to precisely determine sample allele-specific expression

1.2 Protocol for the Detection of LOI Through the Quantification of DNA Methylation Levels

An imprinting center (IC) is a region of DNA that regulates *in cis* the expression of neighboring imprinted genes. ICs are usually characterized by differential DNA methylation, and may be referred to as imprinting control regions (ICRs) or differentially methylated regions (DMRs). Since ICRs are methylated on only one allele in a parent of origin-specific manner, their mean methylation levels are about 50 %. Any deviation from this value, resulting in ICR methylation levels lower or higher than expected [5], is suggestive of an imprinting defect.

Pyrosequencing can accurately quantify the methylation levels at ICRs and therefore identify methylation defects. However, it is very important to define the normal methylation ranges by analyzing a consistent number of healthy subjects.

The analysis requires the initial conversion of genomic DNA using sodium bisulfite, which modifies the unmethylated “C” present in the sequence in “T,” thus translating an epigenetic modification into a difference in the sequence. By Pyrosequencing, both the “C”- and the “T” bases present in the analyzed fragment are sequenced and the methylation value at each CpG site is calculated by the CpG software, which converts the ratio between “C” and “T” incorporation into a methylation percentage.

The feasibility of this technique depends on: (1) recruitment and analysis of an appropriate number of controls; (2) the availability of genomic DNA that will be bisulfite-treated.

2 Materials

2.1 Materials for Detection of LOI by SNP Analysis

Required samples: Genomic DNA and RNA.

2.1.1 Assay Design

1. PCR primer design for the genomic DNA amplification and sequencing.
2. Primer design for cDNA amplification and Pyrosequencing can be made manually or by the PSQ Assay Design software (Qiagen).

2.1.2 Sample Preparation

1. Kit for extraction of genomic DNA (e.g., QIAamp DNA Mini Kit, Qiagen).
2. Reagents for extraction of RNA (e.g., Tri Reagent, Sigma-Aldrich).
3. NanoDrop Spectrophotometer (NanoDrop Technologies Inc, Wilmington DE) or equivalent to quantify nucleic acids.
4. Reagents for reverse transcription (e.g., High Capacity cDNA Reverse Transcription Kit, Life Technologies, which contains both oligod(T) and random primers).

**2.1.3 Genomic DNA
Amplification and Sanger
Sequencing to Identify
Heterozygous SNPs**

1. PCR reagents (Taq Polymerase, dNTPs, Buffer and MgCl_2 , stored at -20°C , e.g., GoTaq polymerase, Promega).
2. Kit for the purification of PCR products (e.g., Ultraclean PCR Clean-up Kit, CABRU).
3. Sequencing reagents (e.g., Sequencing reaction Buffer and Big Dye terminator, Life Technologies).
4. Kit for the purification of the sequencing products (e.g., ZR DNA Clean-up kit, Zymo Research).
5. Thermocycler.
6. Hi-Di Formamide (Life Technologies).
7. Capillary Sequencer: ABI Prism 3100 (Life Technologies) or equivalent.

2.1.4 cDNA Amplification

1. PCR reagents (Taq Polymerase, dNTPs, Buffer, and MgCl_2 , stored at -20°C , e.g., Go Taq Flexi DNA polymerase, Promega).
2. Thermocycler.

**2.1.5 Pyrosequencing
Analyses and Equipment**

(All products indicated below are supplied by Qiagen or Diatech Pharmacogenetics s.r.l. unless otherwise specified).

1. Streptavidin Sepharose HP beads (GE Healthcare, Uppsala, Sweden).
2. Binding buffer, commercially available or prepared as follows: 10 mM Tris-HCl, 2 M NaCl, 1 mM EDTA, 0.1 % Tween 20, pH 7.6.
3. Denaturing solution, commercially available or prepared as follows: 0.2 M NaOH.
4. Washing buffer, commercially available or prepared as follows: 10 mM TrisAcetate, pH 7.6.
5. Annealing buffer, commercially available or prepared as follows: 20 mM TrisAcetate, 2 mM MgAcetate, pH 7.6.
6. High-purity water (MilliQ).
7. 70 % Ethanol.
8. Plate for Pyrosequencing analysis: PSQ 96 low Plate.
9. 96-well PCR plate.
10. Adhesive lid for 96-well PCR plate.
11. Workstation for single strand sequencing template preparation: vacuum preparation tool using the corresponding filter probes, troughs and vacuum pump.
12. Thermomixer or similar (room temperature).
13. Heating device: heating plate or thermoblock.

14. Thermoplate for 96-well plate.
15. Pyro Mark ID instrument equipped with the Pyro Mark ID software.
16. Cartridge for Pyrosequencing reagent dispensation.
17. Pyrosequencing kit (PSQ reagents: enzyme and substrate mix kit).

2.2 Materials for Detection of LOI by Quantification of DNA Methylation Levels

Required samples: Genomic DNA of pathologic cases and controls.

2.2.1 Assay Design

1. Design of PCR and Pyrosequencing primers using the PSQ Assay Design software (Qiagen).

2.2.2 Sample Preparation and Bisulfite Treatment

1. Kit for genomic DNA extraction (e.g., QIAamp DNA Mini Kit, Qiagen).
2. NanoDrop Spectrophotometer (NanoDrop Technologies Inc, Wilmington DE) or equivalent to quantify DNA.
3. DNA bisulfite conversion Kit (EZ DNA methylation Gold Kit, Zymo research), or equivalent.

2.2.3 Genomic DNA Amplification

1. PCR reagents (Hot Start Taq DNA Polymerase, dNTPs—e.g., Go Taq Hot Start Polymerase, Promega—to be stored at -20°C).
2. Thermocycler.

2.2.4 Pyrosequencing Analyses

Reagents and instruments are the same described in Subheading 2.1.5. For this application the Pyro Q-CpG software is required.

3 Methods

3.1 Methods for the Detection of LOI by SNP Analysis

The protocol for the detection of LOI by SNP analysis is organized in different steps, as depicted in Fig. 1 and detailed in the following section.

3.1.1 Assay Design

1. “In silico” Selection of a panel of SNPs located within the coding sequence of the analyzed imprinted region using genomic databases (<http://www.ncbi.nlm.nih.gov/snp/>; <http://genome-euro.ucsc.edu/>).
2. Design PCR primers for SNP genotyping on genomic DNA following standard rules, such as: primer length about 18–24 bases, relatively equal G/C–A/T distribution, melting temperature of 62–65 $^{\circ}\text{C}$ and no dimer and loop formation. The size of amplification product is typically about 200 bp in length.

3. Design primers for the Pyrosequencing assay containing informative SNPs 18–24 bases in length. One primer is biotin-labelled at its 5′ end (biotin allows the immobilization to Sepharose beads during the sample preparation for Pyrosequencing analysis), and the other is unlabeled (*see Note 1*). PCR products should be as short as possible, preferably within 200 base pairs, because the amplification of short fragments is more reliable also for samples in which the quality of DNA is low (i.e., DNA from paraffin-embedded tissues).
4. Design the sequencing primer for Pyrosequencing of the informative SNPs to be about 15–20 bases in length, complementary to the biotinylated PCR strand and unlabeled. The position of the sequencing primer is flexible within five bases from the SNP. It is important to remember that the SNP can be sequenced on either strand, based on the sequence surrounding the SNP (*see Note 2*). As the sequencing reaction is run at 28 °C, it is crucial to check the primer for self-annealing, especially at the 3′-end. The sequence to analyze should be no longer than 100–120 bases, to ensure optimal Pyrosequencing performances and to reduce consumption of the PyroGold reagents.

3.1.2 Sample Preparation

1. Extract genomic DNA using standard procedures such as phenol/chloroform reagents or commercial kits. DNA samples are stored at 4 °C until use.
2. Extract RNA using commercial kits including a DNase treatment. RNA samples are stored at –80 °C until use, to avoid RNA degradation.
3. Synthesize cDNA by reverse transcription of 500 ng of RNA, using commercial kits. Using a mixture of oligod(T) and random primers the reverse transcription of total RNA is ensured. cDNA samples are stored at –20 °C until use.

3.1.3 Genomic DNA Amplification and Sanger Sequencing

This is a preliminary step to identify constitutional informative SNPs eligible for the subsequent evaluation of LOI on cDNA.

1. Use about 0.1–0.2 µg of genomic DNA as substrate for PCR reactions using reagents according to manufacturer's instructions; to obtain a single specific PCR product.
2. A typical PCR protocol for example the amplification of *IGF2* is given in Table 1 using the following thermocycler conditions: Initial denaturation step for 2 min at 94–95 °C followed by 25–35 cycles composed as follows: denaturation step (94–95 °C for 30″–1 min), annealing step (annealing temperature should be set approximately 5 °C below the calculated melting temperature of the primers, for 30″–1 min), extension step (72–74 °C for 1 min for every 1 kb of DNA to be amplified).

Table 1
Setup of a PCR reaction on genomic DNA

Component	Final concentration
PCR buffer	1×
MgCl ₂ solution	2.5 mM
PCR nucleotide mix	0.2 mM each dNTP
Forward PCR primer	0.3 μM
Reverse PCR primer	0.3 μM
DNA polymerase	1.25 units
DNA	0.2 μg
Water	To the final volume (50 μL)

The duration of each step depends on the length of the PCR product. A final extension of 5 min at 72–74 °C is recommended, followed by a refrigeration cycle at 4 °C for several hours after amplification.

- Before sequencing, denature PCR products for 5 min at 96 °C with Hi Di Formamide.
- In the sequencing reaction, the primer is annealed to one of the two template strands and extended in the presence of chemically altered versions of the A, C, G, and T bases, called dideoxynulceotides (ddNTPs). The fluorescently labelled ddNTPs are terminating bases that stop the replication process when incorporated into the growing strand of DNA, resulting in varying lengths of short DNA. These fragments are automatically separated by size, using capillary electrophoresis, from the shortest to the longest one. The whole sequence of the original PCR product is determined by the detection of the fluorescent signal released by the ending ddNTP.
- A typical protocol for the sequencing of PCR products is given in Table 2 using the following thermocycler conditions: Initial denaturation step at 96 °C for 5 min followed by 30 cycles: 95 °C for 30,” 56 °C for 20,” 60 °C for 5 min. At the end, samples can be maintained at 4 °C for several hours.

3.1.4 cDNA Retrotranscription and Amplification

- Perform the RT-PCR according to the manufacturer’s instructions.
- Amplify the region of interest by PCR using the obtained cDNA. A typical PCR protocol is indicated in **step 2** of Subheading 3.1.3.

Table 2
Setup of a sequencing reaction

Component	Amount
PCR product 100–200 bps	1–3 ng
200–500 bps	3–10 ng
500–1,000 bps	5–20 ng
1,000–2,000 bps	10–40 ng
>2,000 bps	20–50 ng
Sequencing reaction buffer	0.75×
Sequencing primer	0.6 μ M
Big dye terminator	1 μ L
Sterile water	To the final volume

3. A prerequisite for a successful Pyrosequencing reaction is the presence of a strong and specific PCR signal. Usually, we use 100 ng of DNA for amplification. In order to obtain it and to avoid residues of the biotinylated primer, 40 or more PCR cycles are recommended.
4. Optionally, verify the PCR products quality by loading a small aliquot (about 3–5 μ L) of the samples on 2 % agarose gel: a single, strong band, with the expected molecular weight should be present, unspecific bands as well as a signal in the negative control must be absent.

3.1.5 Pyrosequencing Analysis

cDNA Pyrosequencing is performed on heterozygous SNPs identified in Subheading 3.1.3.

1. Depending on the amount of PCR product, add 20–40 μ L of each PCR sample to the Pyrosequencing binding mix, composed by 37 μ L of Binding Buffer and 3 μ L of Streptavidin-coated Sepharose beads, into a 96-well PCR. Add HPLC purified water to reach a final volume of 80 μ L, if necessary. Incubate the mixture for 15–30 min on the thermomixer at 1,400 rpm, at room temperature, to allow the binding of the biotinylated PCR products to the Streptavidin-coated Sepharose beads.
2. Purify the single strand sequencing template using the Workstation and exploiting the biotin–streptavidin interaction, following the manufacturer's instruction and as previously reported ([6], *see also* Chapter 7).
3. Release the ssDNA template into a PSQ 96 plate low, containing the Pyrosequencing annealing mix, composed by sequencing primers (0.5–1 μ M) and annealing buffer, to a final volume of 40 μ L for each sample.

4. Denature the template with a thermoplate placed on a thermoblock at 80 °C for 2 min; the subsequent cooling at room temperature allows the primer annealing.
5. Perform the Pyrosequencing reaction using the PyroMark ID instrument. Analyze each sample in duplicate. The PyroMark ID software indicates the amount of PSQ reagents to insert in the cartridge, depending on the “sequence to analyze” and the “dNTPs order dispensation,” previously entered in the software. The “sequence to analyze” should start at the nucleotide upstream the 3′ end of the sequencing primer.
6. Analyze the obtained Pyrograms in the AQ mode. The relative amount of each one allele, corresponding to the amount of each polymorphic nucleotide is indicated as percentage (%).

3.2 Methods for the Detection of LOI by the Quantification of DNA Methylation Levels

The protocol for the detection of LOI by the measurement of the DNA methylation percentage is organized in different steps, as depicted in Fig. 2 and detailed in the following section.

3.2.1 Assay Design

The design of PCR and Pyrosequencing primers is a critical step when working with bisulfite-converted DNA. This issue has been comprehensively discussed in reference [6] and the main points will be briefly described here.

1. Since in the converted DNA molecules the unmethylated “C” are transformed in “T,” sequences frequently show stretches of “T” repeats, which make it difficult to design sequence-specific primers. For this reason, the use of dedicated software (i.e., the PSQ assay design software) is recommended. Take into account that primers with a stable 5′-end and a relatively unstable 3′-end typically work better because they are more stable and specific, and therefore less prone to mispriming.
2. Primers must be specific for the converted sequence and must not anneal to the unconverted DNA (*see Note 3*).
3. In order to avoid degenerated bases in the primer sequence, primers must not contain potentially methylated CpG sites.
4. PCR primers should be 25–30 bases in length, one HPLC purified and biotin-labelled at 5′, the other unlabeled. Sequencing primer should be about 20 bases in length, complementary to the biotinylated PCR strand and unlabeled.
5. The amplification product length should be 150–250 bp in length.
6. Since the sequencing reaction is run at 28 °C, sequencing primers must be checked for the self-annealing, especially at the 3′-end.

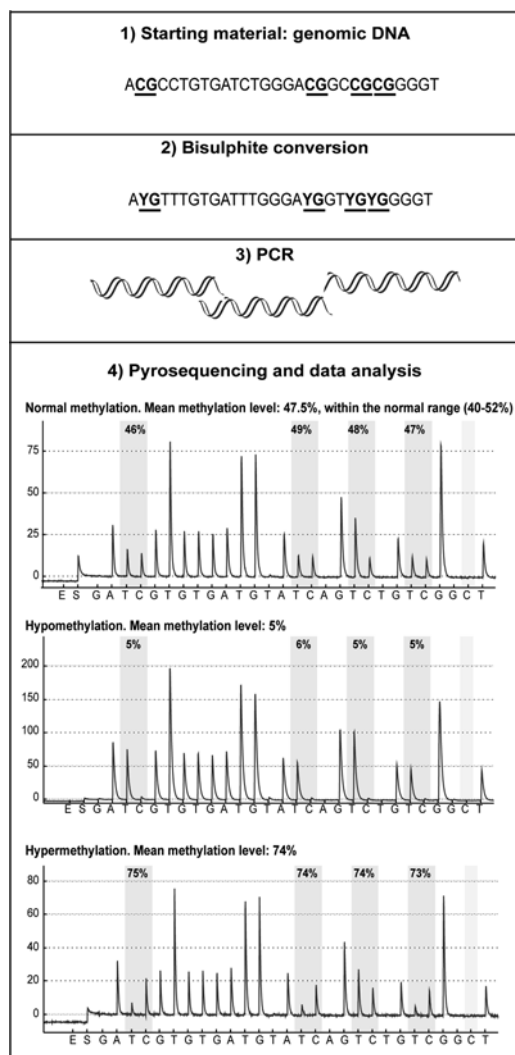


Fig. 2 Schematic representation of the protocol to identify LOI by analyzing DNA methylation. Genomic DNA is (1) extracted and (2) bisulfite converted, in order to differentiate the methylated and unmethylated allele; (3) PCR is used to amplify an Imprinting Control Region; (4) Pyrosequencing experiments showing different methylation levels are displayed. The results are analyzed by the CpG software, which gives the methylation values at each analyzed CpG as percentage

Concerning the Pyrosequencing assay, at least one “bisulfite-treatment control” should be included in the “sequence to analyze” (*see Note 4*).

The sequence to analyze should be no longer than 100–120 bases as described in the previous section. In case of longer sequences to analyze, the use of several sequencing primers is recommended.

3.2.2 Sample Preparation

1. Extract genomic DNA using standard procedures such as phenol/chloroform extraction or commercial kits. Store DNA samples at 4 °C until use.
2. Treat DNA samples with sodium bisulfite using the EZ DNA methylation Gold Kit, following the manufacturer's instructions (*see Note 5*). The modified DNA is stored at -20 °C until use. To ensure reliable results, modified DNA should be used for PCR within 2 months after its conversion.

3.2.3 Amplification of Bisulfite Modified DNA

1. Perform PCR using reagents according to manufacturer's instructions. Notably, in presence of modified DNA, the use of a "Hot-start" Taq polymerase ensures good performances. 2–3 µL of converted DNA are used as template in the PCR reaction.
2. A typical PCR protocol for the amplification of a target region is given in Table 3 using the following thermocycler conditions: Initial denaturation step: 2 min at 94–95 °C followed by 25–35 cycles composed as follows: denaturation step (94–95 °C for 30"–1 min), annealing step (annealing temperature should be set approximately 5 °C below the calculated melting temperature of the primers, for 30"–1 min), extension step (72–74 °C for 1 min for every 1 kb of DNA to be amplified). The duration of each step depends on the length of the PCR product. A final extension of 5 min at 72–74 °C is recommended, followed by a refrigeration cycle at 4 °C for several hours after amplification.
3. To verify that both methylated and unmethylated alleles are amplified with the same efficiency, perform PCR experiments with samples with known methylation levels in parallel. If they are not available, fully methylated or unmethylated synthetic oligonucleotides can be used.

Table 3
Setup of a PCR reaction on bisulfite treated DNA

Component	Final concentration
PCR buffer	1×
MgCl ₂ solution	1.0–4.0 mM
PCR nucleotide mix	0.2 mM each dNTPs
Forward PCR primer	0.1–1 µM
Reverse PCR primer	0.1–1 µM
Hot start DNA polymerase	1.25 units
Modified DNA	About 0.20 µg
Water	To the final volume

4. In order to ensure the maximum amplification efficiency and the complete depletion of the biotinylated primer, perform at least 40 PCR cycles.
5. Optionally, verify the quality of PCR products by loading a small aliquot (about 3–5 μL) of samples on a 2 % agarose gel: a single, strong band, with the expected molecular weight should be present, unspecific bands as well as a signal in the negative control should be absent. PCR products can be stored at $-20\text{ }^{\circ}\text{C}$ for several months or at $+4\text{ }^{\circ}\text{C}$ for several days.

3.2.4 Pyrosequencing Analyses

1. Purify and prepare the single strand sequencing template and perform the Pyrosequencing reaction as described in Subheading 3.1.5.
2. Analyze the Pyrograms using the Pyro Q-CpG software. The ratio of relative amounts of “T” and “C,” at each CpG site, is quantified as percentage that reveals the mean methylation levels of the investigated region.

4 Data Analysis and Applications

4.1 Data Analysis of LOI by SNP Analysis

The raw percentages of allelic contribution calculated by the software can be biased by different elements (e.g., PCR/Pyrosequencing performances), and therefore, a normalization of the data by comparison with reference samples is required. Reference samples are two DNA specimens each homozygous for one of the alleles of the investigated SNP. As example, in Fig. 3 are shown the results obtained for the SNP rs2839702 located on chromosome 11p15.5 at exon 5 of the imprinted gene *H19* (variation: T/G; PCR strand analyzed: reverse A/C) [5].

The standard curve can be obtained as follows:

1. Amplify reference samples by PCR.
2. Mix at least 5 different proportions of the two reference samples and analyze by Pyrosequencing (Fig. 3a: “Expected” % of each sample).
3. Analyze Pyrograms by the PyroMark ID software using the AQ mode. The software quantifies the relative amount of each allele as percentage (Fig. 3a: “Experimental % of each value,” b).
4. Plot the “experimental” % of each allele in a graph in function of its “expected” %. By interpolating the points on the graph it is possible to obtain a straight line, which can be used as standard calibration curve.
5. Calculate the percentage of each allele of the sample to be tested by interpolating the experimental value with the calibration line (Fig. 3c).

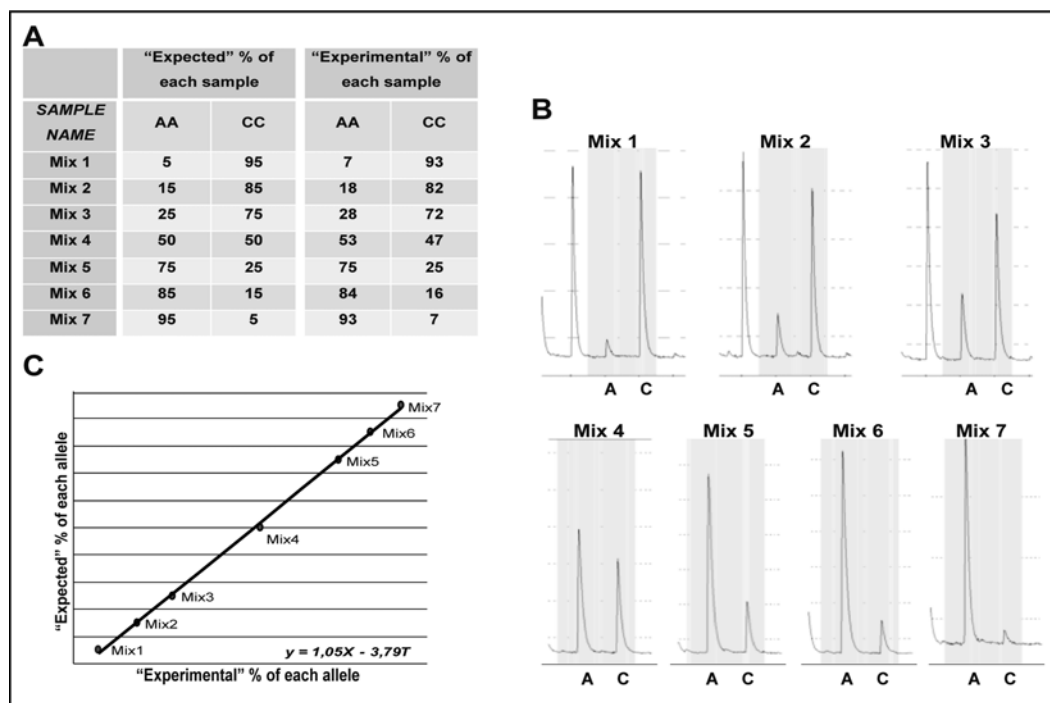


Fig. 3 Method to generate a standard curve for allele-specific expression quantification. In (a) (left column: "Expected" % of each sample), the proportions of PCR products used to generate standard samples are indicated. As specified, seven "MIX" samples were created and analyzed by Pyrosequencing. The allelic expression data obtained by the PyroMark ID instrument (AQ mode), are indicated as "Experimental" % of each sample in (a) (right column) and the corresponding Pyrograms are depicted in (b). The calibration curve, obtained by interpolating the expected "AA" values with the obtained ones is shown in (c). By interpolating the experimental value in the straight line, it is possible to precisely quantify allele-specific expression

4.2 Data Analysis of LOI Through Quantification of DNA Methylation Levels

4.2.1 Definition of Normal Methylation Levels at ICR

The Pyrosequencing of specimens from a sufficiently large number (e.g., 100) of healthy individuals is required to define the range of methylation levels in the general population. In theory, the methylation levels of an ICR in control samples should be closely clustered around the mean value of 50 % (e.g., methylation at the imprinted allele: 100 %; methylation at the non-imprinted allele: 0 %). Our experience suggests that the experimental methylation values, measured by the PyroMark system, can vary within some limits, probably depending on PCR/Pyrosequencing performance and/or physiological variations. For molecular diagnosis of imprinting defects, it is very useful to detect the normal methylation patterns at ICRs, since abnormal methylation can be primary or secondary to other genetic defects such as uniparental disomies (UPDs), ICR microdeletions, or duplications.

1. Analyze about 100 normal subjects to define the normal range.
2. Calculate the normal ranges of methylation variation by analyzing DNA obtained from the same tissue type of the pathologic cases.

3. It is important to run the experiments in duplicate: when two independent experiments of the same sample give methylation levels with discrepant results (>5 %), they should be repeated a third time.

4.2.2 Detection of Methylation Defects at 15q11-q13

Prader–Willi syndrome (PWS; OMIM 176270) and Angelman syndrome (AS; OMIM 105830) are developmental disorders resulting from inappropriate expression of imprinted genes located at 15q11-q13. As reported in Table 4, both PWS and AS are caused by genetic or epigenetic defects in the q11-13 imprinted region of chromosome 15, harboring clusters of genes whose expression is controlled by a bipartite imprinting center (AS/PWS-IC) [7].

Currently, the Pyrosequencing approach allows the identification of methylation defects that can be the result of different genetic/epigenetic mechanisms of PWS and AS (Table 1) [8–10]. Hypermethylation at AS/PWS-IC is sufficient for molecular PWS diagnosis, because the different genetic and epigenetic pathological mechanisms involved in PWS lead to IC hypermethylation (Table 4). By contrast, the molecular diagnosis of AS requires a combination of DNA sequencing, to identify possible *UBE3A* mutations, and methylation analysis at AS/PWS-IC (Table 4).

Table 4

Adapted from GeneReview (<http://www.ncbi.nlm.nih.gov/books/NBK1116/>)

Genetic mechanisms		Occurrence (%)	Methods of detection
PWS	Deletion of 5–6 Mb on the paternal allele	65–75	Karyotype, FISH, aCGH, MS-MLPA, Southern blot Pyrosequencing*
	Chromosomal rearrangement	<1	Karyotype, FISH, Pyrosequencing*
	Maternal UPD	20–30	Microsatellite/SNP analysis, Southern blot, MS-MLPA, Pyrosequencing*
	ID with deletion in the ICRs	0.1	Southern blot, aCGH, Pyrosequencing*
	Epimutation: ID without deletion in the IC	2.5	MS-MLPA, Pyrosequencing
AS	Deletion of 5–7 Mb on the maternal allele	65–75	Karyotype, FISH, aCGH, MS-MLPA, Pyrosequencing*
	Paternal UPD	3–7	Microsatellite/SNP analysis, Southern blot, and MS-MLPA, Pyrosequencing*
	ID with deletion in the ICR	0.5	Southern blot, aCGH, Pyrosequencing*
	ID without deletion in the ICR	2.5	Southern blot, MS-MLPA, Pyrosequencing
	<i>UBE3A</i> mutations	5–11	Sequencing
	“Other”—no identifiable molecular abnormality	10–15	

ID imprinting defect, FISH fluorescent in situ hybridization, aCGH array comparative genomic hybridization, MS-MLPA methylation specific-multiplex ligation-dependent probe amplification. *: in these cases the Pyrosequencing assay can detect the methylation defects due to the genetic mechanisms.

For PWS and AS the assay performed by PyroMark ID instrument targets the 5' CpG island of the *SNURF-SNRPN* (typically referred to as *SNRPN*) locus, located within the AS/PWS-IC that is required for switching to and/or maintenance of the paternal epigenotype. This region is unmethylated on the paternal allele and methylated on the maternal one [11].

These points summarize the possible different pattern of methylation at AS/PWS-IC:

1. Normal individuals have both the methylated (maternally derived) and the unmethylated allele (paternally derived), and therefore, the mean methylation level in controls is about 50 %.
2. Individuals with PWS show hypermethylation that can be associated with: deletions of the paternal allele (65–75 % of PWS cases), maternal UPD (20–30 % of cases), or imprinting defects (2.5 % of cases). In PWS cases the methylation level identified by Pyrosequencing is about 100 %.
3. AS individuals (negative for *UBE3A* mutations) show hypomethylation that can be associated with: deletions of the maternal allele (65–75 % of cases), paternal UPD (3–7 % of cases) or imprinting defects (3 %). The methylation level identified by Pyrosequencing is about 0 %.

4.2.3 Detection of Methylation Defects at 11p15.5

As indicated in Table 5, BWS (OMIM 130650) and SRS (OMIM 180860) are characterized by various epigenetic and/or genetic alterations at the p15.5 imprinted region of chromosome 11. This region contains several genes, whose expression is controlled by two neighboring imprinted domains: ICR1 (Imprinting Control Region 1, regulating *IGF2/H19* genes) and ICR2 (regulating *KCNQ1/CDKN1C* genes). ICR1 is imprinted in the male germ line and operates as an insulator. ICR2 is imprinted in the female germ line and acts as a promoter for the regulatory noncoding RNA *KCNQ1OT1* [12].

Pyrosequencing assays have previously been described to evaluate quantitatively the methylation levels at ICR1 and ICR2 [12]. Since the mechanism originating 11p15.5 methylation defects in SRS and BWS patients is frequently post-zygotic (e.g., UPDs), in the majority of cases the epigenetic alterations are present as mosaicism. For this reason, in order to identify an aberrant methylation pattern even in cases with low-grade mosaicism, it is very important to accurately define the normal methylation range by analyzing peripheral blood lymphocytes of healthy individuals. In presence of such alterations, the Pyrosequencing technology is reliable to detect also minor methylation changes.

It is essential to compare the methylation levels at ICR1 and ICR2 detected in pathological cases with the thresholds defined in controls in order to classify each case as normal or pathological (see Note 6).

Table 5Adapted from reference [13] and from GeneReview (<http://www.ncbi.nlm.nih.gov/books/NBK1116/>)

	Genetic mechanisms	Occurrence (%)	Methods of detection
BWS 11p15.5	Paternal UPD	10–20	Microsatellite/SNP analysis, Southern blot, and MS-MLPA, Pyrosequencing*
	ID at ICR1 on the maternal allele	2–8	Southern blot, MS-MLPA, Pyrosequencing
	ID at ICR2 on the maternal allele	50–60 %	Southern blot, MS-MLPA, Pyrosequencing
	Paternal 11p structural rearrangements	1–2	FISH, aCGH, Pyrosequencing*
	Mutations of the maternal <i>CDKN1C</i> allele	5–10 (sporadic) 40 (AD traits)	sequencing
SRS 11p15.5-related	ID at ICR1 on the paternal allele	35–60	Southern blot, MS-MLPA, Pyrosequencing
	Maternal duplications	4	Karyotype, FISH, aCGH, Pyrosequencing*
SRS Chr7-related	UPD (maternal)	7–10	Microsatellite/SNP analysis, Southern blot, and MS-MLPA, Pyrosequencing*
	Deletion/duplication	Rare	Karyotype, FISH, aCGH, Pyrosequencing*

UPD uniparental disomy, *ICR* imprinting control region, *ID* imprinting defects, *AD* autosomal dominant, *FISH* fluorescent in situ hybridization, *aCGH* array comparative genomic hybridization, *MS-MLPA* methylation specific-multiplex ligation-dependent probe amplification. *: in these cases the Pyrosequencing assay can detect the methylation defects due to the genetic mechanisms.

5 Notes

1. As free biotin competes for binding to streptavidin with the biotinylated PCR product, purification of the biotinylated primer by HPLC is recommended (free biotin lowers the signal level of the Pyrosequencing reaction).
2. The interpretation of the SNP could be difficult when one or both of the polymorphic bases will form a homopolymer with the adjacent bases. It is preferable, in these cases, to design sequencing primer oriented with the 3'-end overlapping the polymeric stretch.
3. PCR primers should be complementary to regions containing four or more C outside a CpG site because these bases discriminate between converted and non-converted DNA independent of the methylation status. In addition, it is possible to test PCR primers specificity by running a PCR of a non-converted DNA sample: PCR product should be absent.

4. The bisulfite control is a “C” not followed by a “G,” and therefore, the C is not methylated and it is converted to “T” by bisulfite. Only “T” should be incorporated in the sequence and the incorporation of “C” at this position indicates an incomplete bisulfite conversion.
5. Even if in the EZ methylation Kit protocol the elution volume of converted DNA is 10 μ L, it can be increased to 30 μ L without lowering the quality of DNA. Then, 2–3 μ L can be used for PCR.
6. At present, Pyrosequencing has a high level of accuracy to study methylation variation, especially in cases with low-level mosaicism, compared with semiquantitative (southern blotting and methylation-sensitive multiplex ligation probe analysis) and qualitative (methylation-specific PCR) approaches.

References

1. Feinberg AP, Cui H, Ohlsson R (2002) DNA methylation and genomic imprinting: insights from cancer into epigenetic mechanisms. *Sem Cancer Biol* 12:389–398
2. Rainier S, Johnson LA, Dobry CJ et al (1993) Relaxation of imprinted genes in human cancer. *Nature* 362:747–749
3. Djiogue S, Nwabo Kamdje AH, Vecchio L et al (2013) Insulin resistance and cancer: the role of insulin and IGFs. *Endocr Relat Cancer* 20:R1–R17
4. Delaval K, Wagschal A, Feil R (2006) Epigenetic deregulation of imprinting in congenital diseases of aberrant growth. *Bioessays* 28:453–459
5. Tabano S, Colapietro P, Cetin I et al (2010) Epigenetic modulation of the IGF2/H19 imprinted domain in human embryonic and extra-embryonic compartments and its possible role in fetal growth restriction. *Epigenetics* 5:313–324
6. Dejeux E, El abdalaoui H, Gut IG et al (2009) Identification and quantification of differentially methylated loci by the pyrosequencing technology. *Methods Mol Biol* 507: 189–205
7. Kantor B, Shemer R, Razin A (2006) The Prader-Willi/Angelman imprinted domain and its control center. *Cytogenet Genome Res* 113:300–305
8. Glenn CC, Saitoh S, Jong MT et al (1996) Gene structure, DNA methylation, and imprinted expression of the human SNRPN gene. *Am J Hum Genet* 58:335–346
9. Kubota T, Sutcliffe JS, Aradhya S et al (1996) Validation studies of SNRPN methylation as a diagnostic test for Prader-Willi syndrome. *Am J Med Genet* 66:77–80
10. Glenn CC, Driscoll DJ, Yang TP et al (1997) Genomic imprinting: potential function and mechanisms revealed by the Prader-Willi and Angelman syndromes. *Mol Hum Reprod* 3:321–332
11. Horsthemke B, Wagstaff J (2008) Mechanisms of imprinting of the Prader-Willi/Angelman region. *Am J Med Genet A* 146A:2041–2052
12. Calvello M, Tabano S, Colapietro P et al (2013) Quantitative DNA methylation analysis improves epigenotype-phenotype correlations in Beckwith-Wiedemann syndrome. *Epigenetics* 8:1053–1060
13. Eggermann T, Eggermann K, Schönherr N (2008) Growth retardation versus overgrowth: Silver-Russell syndrome is genetically opposite to Beckwith-Wiedemann syndrome. *Trends Genet* 24:195–204

Chapter 19

Analysis of DNA Methylation Patterns in Single Blastocysts by Pyrosequencing®

John Huntriss, Kathryn Woodfine, Joanna E. Huddleston, Adele Murrell, and Helen M. Picton

Abstract

Extensive epigenetic reprogramming occurs during mammalian gametogenesis and preimplantation development. DNA methylation patterns that are laid down during these stages are essential for subsequent normal foetal development. The requirement for more precise assessment of the epigenetic programming of in vitro-derived human preimplantation embryo has become of paramount importance following the identification of epigenetic diseases that are associated with assisted reproduction and/or infertility. Such techniques are also useful and applicable to experimental reproductive biology. In order to expand our knowledge of epigenetic marks, including DNA methylation, during mammalian reproduction and early development, it is necessary to test new and sufficiently sensitive protocols. There are, however, unique challenges to obtain DNA methylation data from the small cell numbers that are present in the preimplantation embryo. In this protocol, we describe the successful application of Pyrosequencing® to yield quantitative DNA methylation data over several CpG sites at differentially methylated regions (DMRs) at imprinted loci in single blastocysts, in this case, human blastocysts. Future developments of the protocol will allow DNA methylation analysis of a more extensive panel of genes for each embryo and at the same time, since the protocol allows for the extraction of mRNA from the embryo, the comparison between DNA methylation and gene expression.

Key words Blastocyst, Preimplantation, Pyrosequencing®, DNA methylation, Epigenetic, Imprinted gene, Assisted reproduction, ART, IVF

1 Introduction

Epigenetic information is highly dynamic during mammalian gametogenesis and preimplantation development [1]. The analysis of gene methylation in single mammalian preimplantation embryos and oocytes has to date relied upon bisulphite treatment of an agarose bead-embedded oocyte/embryo, followed by PCR for the region of interest and then cloning into a vector and clone sequencing. That method has provided a robust protocol that has been used over many key studies [2–7]. However, it is also important to

explore new methods of DNA methylation analysis, particularly quantitative analysis of methylation and/or to look towards whole genome analysis. These approaches are necessary in order to expand our knowledge of epigenetic regulation during reproduction and early development.

The need for performing single embryo analysis is particularly relevant to the study of human blastocysts for several reasons. Human embryos are only available for research in small numbers due to (1) the prioritisation of their use in clinical treatment as part of the assisted reproduction technology (ART) treatment programme, (2) the requirements for patient consent for use in research, (3) the requirements for appropriate licensing and ethical approval of performing analyses of this nature. That only a restricted number of embryos can be generated per patient per cycle further restricts their availability for research purposes.

Therefore, any protocols developed to facilitate the in-depth molecular-genetic/epigenetic evaluation of human blastocyst-staged preimplantation embryos must be geared towards single embryo analysis rather than being conducted on pooled samples of embryos. The advantages of this strategy are that the epigenetic properties of each individual embryo could potentially be compared against routinely captured data including embryo morphology and fragmentation, rate of development, metabolic profile, or the suspected cause of infertility. In that way, a much more comprehensive profile of the epigenetic programming that is related to preimplantation human embryo development can be built. Techniques that work at single blastocyst level may also be advantageous in animal studies as a means of restricting costs and reducing animal usage by down-scaling experiments and by avoiding incorrect data interpretation, which is inherent when pooling due to effects caused by interindividual variation.

The intention of the protocol and experiments described here include (1) to achieve proof-of-principal that DNA methylation analysis by Pyrosequencing[®] works on single human blastocyst-staged preimplantation embryos, (2) to establish which imprinted genes have detectable methylation imprints at this stage of human development, and (3) to investigate whether Pyrosequencing-based DNA methylation analysis is a technique that could be used to improve our understanding about the epigenetic health of in vitro derived human embryos.

With reference to this third point, there is evidence that ART may cause epigenetic disease, although factors relating to the underlying infertility may also be involved [8–10]. Increased incidence of the imprinting disorder Beckwith–Wiedemann syndrome (BWS) has been described following the use of ART [11–19]. Where tested, the epimutation in the majority of these instances involved abnormal methylation at KvDMR1, a maternally methylated imprinting control element on chromosome 11p15.5 located

on the promoter of the *KCNQ1OT1* gene [20]. Furthermore, an association between ART and the imprinting disorder Angelman syndrome has been reported [16, 21–23], with hypomethylation at the *SNRPN* imprinting control region observed. An increased frequency of Silver–Russell syndrome associated with ART and corresponding abnormal methylation at a number of gametic DMRs of imprinted genes has been reported more recently [19]. Epigenetic defects are also associated with infertility and aberrant methylation of imprinted genes has been observed in spermatozoa from men with a range of defects in spermatogenesis [24–27].

It was considered that the quantitative nature of DNA methylation analysis by Pyrosequencing would offer the best technique for accurately assessing methylation at differentially methylated regions (DMRs) of imprinted genes in blastocysts. The PCR protocol for Pyrosequencing of the human imprinted gene differentially methylated regions (DMRs) was originally described by Woodfine et al. [28], and is particularly suitable to meet the high level of sensitivity that is essential when analysing individual blastocysts. Importantly, the protocol incorporates two rounds of PCR for the region of interest. Therefore, following bisulphite treatment of embryonic genomic DNA derived from the entire embryo, the first PCR round is performed with the gene specific primers (with the reverse primer incorporating an M13 sequence) and the products are subsequently amplified in the second round with a gene-specific forward primer and a common biotinylated reverse primer (containing an M13 site). This arrangement was originally designed to reduce assay costs, increase sensitivity, and allow high throughput amplification of the second round [28]. An overview of the protocol is given in Fig. 1. The accurate assessment of methylation in single preimplantation embryos by use of this Pyrosequencing protocol is therefore a promising development towards precise assessment of the epigenetic status of the human blastocyst-staged preimplantation embryo in response to the in vitro culture environment, ART regimen and possible infertility-associated epimutations.

2 Materials

1. Lysis buffer for embryos: Dynabeads® mRNA DIRECT™ Micro Kit (Life Technologies).
2. DNA isolation from embryos: Qiagen QIAamp DNA Micro kits.
3. Control DNA samples of adult human kidney and blood were obtained from AMS Biotechnology.
4. Bisulphite conversion: EZ DNA Methylation-Gold kit (Zymo Research Corporation, CA, USA).

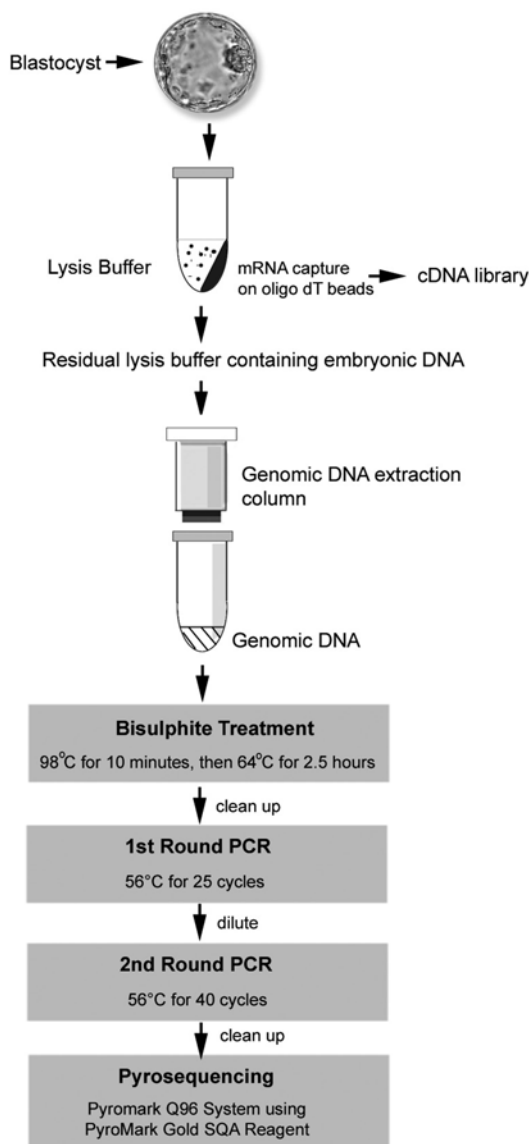


Fig. 1 Overview of the protocol used for obtaining quantitative methylation data by Pyrosequencing® from single blastocysts. The protocol allows for optional isolation of gDNA and cDNA

5. PCR primer (desalted), *see* Subheading 3.3 and Table 1.
6. Biotinylated M13 reverse primer.
7. ThermoStart ABGene PCR Master Mix (Fisher, Loughborough, UK).
8. PCR product cleanup for Pyrosequencing: Streptavidin Sepharose high performance beads (GE Healthcare, Buckinghamshire, UK) and Binding Buffer (Qiagen, Hilden, Germany).

Table 1
Adapted from ref. [28]

Gene assay	Chromosome	Forward primer	Reverse primer	Sequencing primer
<i>DIRAS3</i> (2)	1	TGGATTAGTTTTTTAGATTGTTGTAGATGT	CCCCAAAAAACTCACTCCTCC	GGTTAGTTTTTTATAGTTGGT
<i>DIRAS3</i> (1)	1	AACTCCCTCTAACTTCTTCTCCTTACCT	GTTTATAGGGTATTGTTGTTAAAAATTGT	CTTCTCCCTTACCTAT
<i>GRB10</i> (g)	7	CTCTCCAAATACTCAAATAAACTC	GGTAGGGGTTTGTGTAGTTTG	CCAAAACTCAAAATAAACTCC
<i>KvDMR</i>	11	AGGGAAGTTTTTAGGGTGTGAATTTTTTAGAG	CCAAACCAACCCACCTAACAAAAAAC	TGGTAATGTTTGGTATTT
<i>RBI</i>	13	GGTAGGGTAGTTTGGAAATGTTTAAG	AACCACAAACCCCTTACCC	AGTTTGGGAAATGTTTAAGAT
<i>SNRPN</i>	15	CCTACCTCCCAACCACTTCCT	GGGAGATTGGGGGGGAATT	CCCAACCACTTCCTA

9. Sample preparation for sequencing: PyroMark® vacuum tool and associated buffers (Qiagen).
10. Hardware: Samples were analysed on a PyroMark® Q96 machine (Qiagen).
11. Reagents for pyrosequencing: PyroMar®k Gold Q96 SQA Reagents (Qiagen).
12. Analysis: Pyrograms® were analyzed using the PyroQ CpG software and graphs were generated in Graphpad Prism 5.0.

3 Methods

3.1 Genomic DNA Isolation from Blastocysts

In this protocol, lysed human blastocyst-staged preimplantation embryos that were consented for research purposes as described elsewhere [29] were routinely stored at -80°C in a minimal volume ($<20\text{ }\mu\text{L}$) of Dynal lysis buffer, a buffer that is supplied for the Oligo-dT Dynabeads micro kit. The rationale for the use of this buffer was that it allowed the isolation of both the genomic DNA and also the mRNA from a single embryo, with the potential to allow comparison of the transcriptome and the epigenome of a single embryo, and thus maximizing the information gained from these precious research samples. The mRNA, as isolated from embryos by Oligo dT Dynabeads was subsequently amplified using a cDNA amplification method, as described by Huntriss et al. [30].

1. Isolate embryonic genomic DNA using the Qiagen DNA micro kit from the lysed embryo that remained in approximately $20\text{ }\mu\text{L}$ of residual Dynal lysis buffer after the oligo-dT beads and bound mRNA had been removed (*see Note 1*). DNA was isolated essentially according to the manufacturer's protocol (Qiagen).
2. Add $200\text{ }\mu\text{L}$ ATL buffer (*see Note 2*), $10\text{ }\mu\text{L}$ proteinase K, incubate for 45 min at 56°C .
3. Add $200\text{ }\mu\text{L}$ AL buffer and $1\text{ }\mu\text{L}$ carrier RNA (*see Note 3*), mix by vortexing, and incubate for 10 min at 56°C .
4. Add $90\text{ }\mu\text{L}$ 100 % ethanol, mix by vortexing, and incubate at room temperature for 3 min.
5. Add the entire mix to a QIAamp MinElute column, spin at $4,450\times g$.
6. Wash with AW1 buffer (with ethanol added according to kit instructions).
7. Spin at $4,450\times g$ for 1 min.
8. Wash with AW2 buffer (with ethanol added according to kit instructions).
9. Spin at $4,450\times g$ for 1 min and discard eluate.

10. Place the MinElute column in a clean collection tube.
11. Spin for 3 min at $12,650\times g$ to dry the membrane.
12. Add 20 μL distilled water to column membrane and incubate for 1 min at room temperature.
13. Spin for 1 min at $12,650\times g$ to elute genomic DNA.
14. Genomic DNA was immediately vacuum desiccated and stored at -80°C .

3.2 Bisulphite Treatment of Blastocyst Genomic DNA

Bisulphite treatment of embryonic genomic DNA was performed using the EZ DNA Methylation-Gold kit essentially according to the manufacturer's instructions.

1. Reconstitute blastocyst genomic DNA pellets by the addition of 20 μL distilled water.
2. Add to 20 μL genomic DNA (the entire embryonic DNA sample), 130 μL of the freshly prepared CT Conversion Reagent in a PCR tube.
3. Run a positive control of adult human kidney DNA, or blood DNA, treated identically to the embryonic DNA, alongside all assays.
4. Incubate the mixture for bisulphite conversion for 10 min at 98°C , and then for 2.5 h at 64°C (*see Note 4*).
5. Add 600 μL of M-Binding Buffer to a Zymo-Spin™ IC Column with the spin column placed in a collection tube.
6. Load the converted embryonic DNA/ CT Conversion Reagent mix onto the Zymo-Spin™ IC Column containing the M-Binding Buffer and mix by inversion.
7. Centrifuge for 30 s at $12,650\times g$ and discard the column flow-through.
8. Wash with 100 μL M-Wash Buffer. Centrifuge for 30 s at $12,650\times g$.
9. Add 200 μL of M-Desulphonation Buffer to the Zymo-Spin™ IC Column. Incubate at room temperature for 20 min.
10. Centrifuge for 30 s at $12,650\times g$.
11. Wash twice with 200 μL M-Wash Buffer. Centrifuge each time for 30 s at $12,650\times g$.
12. Elute the converted DNA from the column by adding 25 μL of water and spinning at $12,650\times g$ into a clean collection tube.

3.3 First Round PCR

The PCR assays for the human imprinted genes that were used on human blastocyst DNA were designed, optimized, and extensively tested as previously described [28]. The Pyrosequencing assays that were successful on single blastocysts included *RBI*, *SNRPN*,

KvDMR1, *GRB10*, as described previously [29]. In addition, two DMRs within *DIRAS3* were successfully tested in pooled blastocyst samples (unpublished results). PCR primers for KvDMR1 were designed using MethPrimer [31] or the PyroMark Assay design software 2.0 (Qiagen) for the remaining assays (*RBI*, *SNRPN*, *GRB10*, *DIRAS3*).

1. First Round PCR on blastocyst DNA was performed in a total volume of 10 μ L in a 96-well PCR plate.
2. Due to the limiting amount of DNA available, the required input of bisulphite converted embryonic DNA should be empirically determined for each assay and could range from 0.5 to 4.0 μ L of DNA as eluted from the clean-up column. Added to this were 0.25 μ M forward primer, 0.25 μ M reverse primer, 5 μ L ThermoStart ABGene PCR MasterMix (Fisher, Loughborough, UK), and distilled water up to 10 μ L. Primers are detailed in Table 1.
3. PCR was performed after an initial 15 min denaturation step at 95 °C with 25 cycles of 95 °C for 30 s, 56 °C for 30 s, 72 °C for 1 min, followed by final extension step at 72 °C for 5 min (*see* **Note 5**).
4. For control human DNA samples that were run alongside the embryonic DNA, the PCR assay was performed using 10 ng of bisulphite converted DNA.

3.4 Second Round PCR

1. Dilute the PCR products from the first round reaction 1:4 to 1:6 times with distilled water for each assay.
2. The volume of diluted second round reaction that was added to the PCR mixture ranged from 0.5 to 6.375 μ L. Add 0.25 μ M forward primer, 0.25 μ M reverse primer, 7.5 μ L ThermoStart ABGene PCR MasterMix, and distilled water up to a 15 μ L final volume.
3. In the second PCR round, the gene-specific forward primer was used in conjunction with a common biotinylated *M13* reverse primer 5' Biotin-CGCCAGGGTTTTCCCAGTCACGAC 3'.
4. PCR was performed after an initial 15 min denaturation step at 95 °C with 40 cycles of 95 °C for 30 s, 56 °C for 30 s, 72 °C for 1 min, followed by final extension step at 72 °C for 5 min.
5. Run a small volume (2–3 μ L) of the second round PCR reaction on an agarose gel to establish whether the PCR for each tested gene was successful and whether the PCR produced a single band.

3.5 Cleanup/Binding

1. For a full 96-well plate, add 240 μ L of Streptavidin Sepharose high performance beads to 4,560 μ L binding buffer and 3,600 μ L distilled water.

2. Add 70 μL of the above mix and 10 μL of second round PCR product to each well of a microtiter plate.
3. Incubate plates for 20 min with shaking.
4. Wash beads with the bound second round PCR product and harvest using the PyroMark Q96 Vacuum Workstation and associated buffers.

3.6 Pyrosequencing Analysis

1. For each sample, add 11 μL PyroMark Annealing Buffer to 1 μL (0.25 μM) gene-specific forward inner sequencing primer [28] and distribute into a Pyrosequencing plate.
2. Deposit the washed, dried beads from the cleanup/binding stage into the 12 μL mixture of Annealing Buffer/sequencing primer on the Pyrosequencing plate by agitation and denature the samples by heating for 3 min at 80 °C on a PCR machine mounted hotplate.
3. Perform Pyrosequencing on a PyroMark Q96 System using PyroMark Gold SQA Reagent kits (Qiagen) mixing 232 μL enzyme mix, 232 μL substrate mix and typically the following volume of nucleotides: A (67 μL), C (57 μL), G (84 μL), T (94 μL).
4. Use the Pyro Q-CpG Software (Biotage, Uppsala, Sweden) for quantifying methylation. Examples of Pyrograms for *SNRPN*, *KvDMR1*, and *RBI* as obtained from single human blastocysts are shown in Fig. 2.

4 Notes

1. It is possible that the use of an alternative embryo lysis buffer, and/or the direct isolation of DNA without prior extraction of mRNA may increase DNA stability and therefore increase the number of assays that could be achieved per single embryo. If transcriptome analysis is a requirement, then there are new products for DNA and RNA co-isolation that do not use oligo-dT and therefore could presumably widen the repertoire of RNA species that can be analysed in relation to the methylation data.
2. Using 200 μL ATL buffer and the subsequent addition of 200 μL AL buffer per single blastocyst is probably in excess of requirements and this volume could probably be reduced.
3. The addition of carrier RNA is recommended in the Qiagen DNA micro kit for increasing yields with limiting amounts of DNA.
4. The use of commercially available sample storage buffers and reagents that are optimized to preserve DNA during the bisulphite conversion procedure could also assist in maximizing the number of assays that can be performed per embryo.

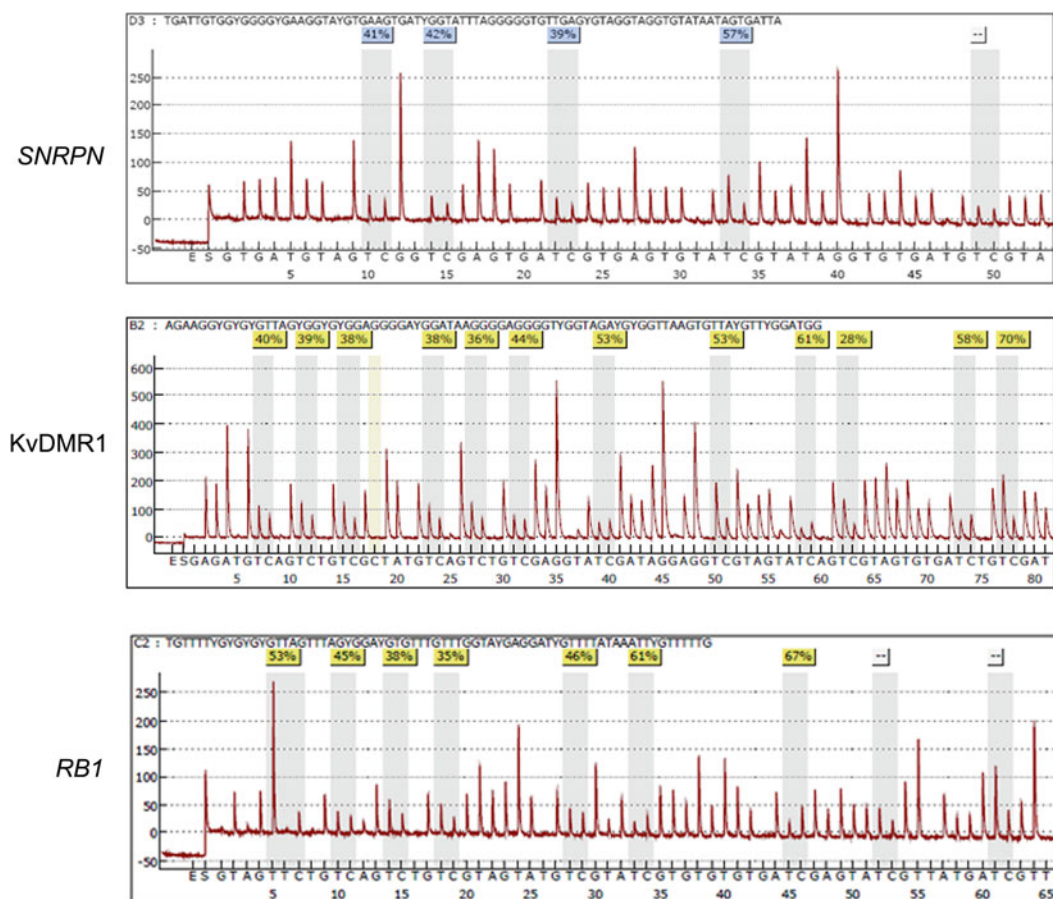


Fig. 2 Example Pyrograms for the DMRs of *SNRPN*, *KvDMR1*, and *RB1* as obtained from three human blastocysts. Adapted from Huntriss J, Woodfine K, Huddleston JE, Murrell A, Rutherford AJ, Elder K, Khan AA, Hemmings K, Picton H. *Fertil Steril*. 2011; 95(8):2564-7, reproduced with permission

5. Of particular note, we observed that it is preferable to perform PCR and Pyrosequencing as soon as possible after bisulphite conversion. This is especially relevant due to the limiting amount of DNA, since any further losses by DNA degradation need to be avoided. For example, storage of residual converted DNA at -80°C for 1 week was sufficient to cause total assay failure.

5 Additional Comments

Multiplexing of PCR assays may represent a way of increasing the number of assays that can be obtained per embryo. Whole bisulphitome amplification of the bisulphite converted DNA, if controlled adequately, may also significantly increase the number of assays that can be performed per blastocyst. It is likely that the

current protocol would need to be adapted in order to generate methylation data from cleavage-staged embryos or indeed single oocytes. Agarose bead embedding of single oocytes and the avoidance of spin columns during the DNA extraction phases and the bisulphite cleanup may be key points in an adapted protocol.

Acknowledgments

Supported by MRC (grant no. G0701388), Ref 06/26 and the Association for International Cancer Research (grant no. 06-0589).

References

1. Morgan HD, Santos F, Green K et al (2005) Epigenetic reprogramming in mammals. *Hum Mol Genet* 14:R47–R58
2. Warnecke PM, Mann JR, Frommer M et al (1998) Bisulphite sequencing in preimplantation embryos: DNA methylation profile of the upstream region of the mouse imprinted H19 gene. *Genomics* 51:182–190
3. Hanel ML, Wevrick R (2001) Establishment and maintenance of DNA methylation patterns in mouse Ndn: implications for maintenance of imprinting in target genes of the imprinting center. *Mol Cell Biol* 21:2384–2392
4. Olek A, Walter J (1997) The pre-implantation ontogeny of the H19 methylation imprint. *Nat Genet* 17:275–276
5. Geuns E, De Rycke M, Van Steirteghem A et al (2003) Methylation imprints of the imprint control region of the SNRPN-gene in human gametes and preimplantation embryos. *Hum Mol Genet* 12:2873–2879
6. Geuns E, Hilven P, Van Steirteghem A et al (2007) Methylation analysis of KvDMR1 in human oocytes. *J Med Genet* 44:144–147
7. Sato A, Otsu E, Negishi H et al (2007) Aberrant DNA methylation of imprinted loci in superovulated oocytes. *Hum Reprod* 22:26–35
8. Chiba H, Hiura H, Okae H et al (2013) DNA methylation errors in imprinting disorders and assisted reproductive technology. *Pediatr Int* 55:542–549
9. Odom LN, Segars J (2010) Imprinting disorders and assisted reproductive technology. *Curr Opin Endocrinol Diabetes Obes* 17:517–522
10. Huntriss J, Picton HM (2008) Epigenetic consequences of assisted reproduction and infertility on the human preimplantation embryo. *Hum Fertil (Camb)* 11:85–94
11. Maher ER, Brueton LA, Bowdin SC et al (2003) Beckwith-Wiedemann syndrome and assisted reproduction technology (ART). *J Med Genet* 40:62–64
12. DeBaun MR, Niemitz EL, Feinberg AP (2003) Association of in vitro fertilization with Beckwith-Wiedemann syndrome and epigenetic alteration of LIT1 and H19. *Am J Hum Genet* 72:156–160
13. Gicquel C, Gaston V, Mandelbaum J et al (2003) In vitro fertilization may increase the risk of Beckwith-Wiedemann syndrome related to the abnormal imprinting of the KCN1OT gene. *Am J Hum Genet* 72:1338–1341
14. Halliday J, Oke K, Breheny S et al (2004) Beckwith-Wiedemann syndrome and IVF: a case-control study. *Am J Hum Genet* 75:526–528
15. Chang AS, Moley KH, Wangler M et al (2005) Association between Beckwith-Wiedemann syndrome and assisted reproductive technology: a case series of 19 patients. *Fertil Steril* 83:349–354
16. Sutcliffe AG, Peters CJ, Bowdin S et al (2006) Assisted reproductive therapies and imprinting disorders: a preliminary British survey. *Hum Reprod* 21:1009–1111
17. Bowdin S, Allen C, Kirby G et al (2007) A survey of assisted reproductive technology births and imprinting disorders. *Hum Reprod* 22:3237–3240
18. Doornbos ME, Maas SM, McDonnell J et al (2007) Infertility, assisted reproduction technologies and imprinting disturbances: a Dutch study. *Hum Reprod* 22:2476–2480
19. Hiura H, Okae H, Miyauchi N et al (2012) Characterization of DNA methylation errors in patients with imprinting disorders conceived by assisted reproduction technologies. *Hum Reprod* 27:2541–2548

20. Lim D, Bowdin SC, Tee L et al (2009) Clinical and molecular genetic features of Beckwith-Wiedemann syndrome associated with assisted reproductive technologies. *Hum Reprod* 24: 741–747
21. Cox GF, Burger J, Lip V et al (2002) Intracytoplasmic sperm injection may increase the risk of imprinting defects. *Am J Hum Genet* 71:162–164
22. Ørstavik KH, Eiklid K, van der Hagen CB et al (2003) Another case of imprinting defect in a girl with Angelman syndrome who was conceived by intracytoplasmic sperm injection. *Am J Hum Genet* 72:218–219
23. Ludwig M, Katalinic A, Gross S et al (2005) Increased prevalence of imprinting defects in patients with Angelman syndrome born to subfertile couples. *J Med Genet* 42:289–291
24. Marques CJ, Costa P, Vaz B et al (2008) Abnormal methylation of imprinted genes in human sperm is associated with oligozoospermia. *Mol Hum Reprod* 14:67–74
25. Hammoud SS, Purwar J, Pflueger C et al (2010) Alterations in sperm DNA methylation patterns at imprinted loci in two classes of infertility. *Fertil Steril* 94:1728–1733
26. Boissonnas CC, Abdalaoui HE, Haelewyn V et al (2010) Specific epigenetic alterations of IGF2-H19 locus in spermatozoa from infertile men. *Eur J Hum Genet* 18:73–80
27. Poplinski A, Tüttelmann F, Kanber D et al (2010) Idiopathic male infertility is strongly associated with aberrant methylation of MEST and IGF2/H19 ICR1. *Int J Androl* 33: 642–649
28. Woodfine K, Huddleston JE, Murrell A (2011) Quantitative analysis of DNA methylation at all human imprinted regions reveals preservation of epigenetic stability in adult somatic tissue. *Epigenetics Chromatin* 4:1
29. Huntriss J, Woodfine K, Huddleston JE et al (2011) Quantitative analysis of DNA methylation of imprinted genes in single human blastocysts by pyrosequencing. *Fertil Steril* 95: 2564–2567
30. Huntriss JD, Hemmings KE, Hinkins M et al (2013) Variable imprinting of the MEST gene in human preimplantation embryos. *Eur J Hum Genet* 21:40–47
31. Li LC, Dahiya R (2002) MethPrimer: designing primers for methylation PCRs. *Bioinformatics* 18:1427–1431

Chapter 20

Allele-Specific DNA Methylation Detection by Pyrosequencing®

Lasse Sommer Kristensen, Jens Vilstrup Johansen,
and Kirsten Grønbæk

Abstract

DNA methylation is an epigenetic modification that plays important roles in healthy as well as diseased cells, by influencing the transcription of genes. In spite the fact that human somatic cells are diploid, most of the currently available methods for the study of DNA methylation do not provide information on the methylation status of individual alleles of genes. This information may be of importance in many situations. In particular, in cancer both alleles of tumour suppressor genes generally need to be inactivated for a phenotypic effect to be observed. Here, we present a simple and cost-effective protocol for allele-specific DNA methylation detection based on Pyrosequencing® of methylation-specific PCR (MSP) products including a single nucleotide polymorphism (SNP) within the amplicon.

Key words Epigenetics, DNA methylation, CpG islands, Allele-specific methylation, Allelic methylation, Methylation-specific PCR, Bisulfite Pyrosequencing®

1 Introduction

In recent years it has become increasingly apparent that DNA methylation, in concert with histone modifications, plays important roles in gene regulation in mammalian cells. This epigenetic mark is faithfully copied through mitotic cell divisions by the maintenance methyltransferase DNMT1, but is generally not inherited through the germline [1]. During early human development the epigenome is reset by a global erasure of methylation marks, which are subsequently re-established by the de novo DNA methyltransferases DNMT3A and DNMT3B in the developing embryo [2]. Common for mammalian DNMTs is that they transfer methyl groups to cytosines in a CpG context.

The CpG dinucleotides are unevenly distributed in the human genome. Most genomic regions are more or less depleted of CpG sites while others have a very high CpG density and are referred to as CpG islands. These islands are often found near the transcription

start site of genes and generally harbor unmethylated CpG sites in healthy cells. Indeed, 60–70 % of annotated gene promoters are associated with one or more CpG islands [3]. By contrast, CpG sites found outside CpG islands are mainly methylated, and may play a role in maintaining genome stability, e.g., by silencing of transposable elements [4]. DNA methylation also plays a critical role in cellular differentiation and development [5, 6], X-chromosome inactivation [7], imprinting [8], and allele-specific gene expression [9].

The possibility for fine-tuning gene expression in different cells and tissues through modulation of the epigenetic landscape is vital for the existence of complex organisms, but may also be regarded as an *Achilles' heel* as many different multifactorial human diseases have a strong underlying epigenetic component [10–12]. In particular, cancer cells are able to alter the expression of hundreds of genes, in part, by altering methylation patterns of promoter CpG islands [13]. While this is valuable knowledge from an academic viewpoint, it is also likely that the detection of DNA methylation will have a profound impact on the clinical management of cancer in the future. Many loci, which become hypermethylated during cancer development, are important candidates as specific biomarkers for early cancer detection, diagnosis, and prognosis [14], and unlike mutations, the aberrant epigenetic states of cancer cells are potentially reversible. Several drugs targeting proteins involved in the modification of the epigenome have been developed and are currently used as standard treatment or being tested in clinical trials. In particular, DNMT inhibitors have already been approved for the treatment of myelodysplastic syndrome (MDS) [15] and AML with low blast counts [16].

Many different methods have been developed for the detection of DNA methylation at single loci [17, 18]. The read-out provided by these methods varies significantly with respect to throughput, analytical sensitivity and specificity, quantitative accuracy, allelic methylation information, etc. [18]. Most PCR-based methods, including Pyrosequencing® [19], rely on a prior sodium bisulfite conversion of the DNA in order to preserve the DNA methylation information. This information is otherwise lost during PCR as no polymerase exists, which is capable of distinguishing cytosine and 5-methylcytosine. By treating the DNA with sodium bisulfite unmethylated cytosines become converted into uracil and subsequently amplified as thymine, while methylated cytosines remain as cytosine. Many different commercially available kits exist, which generally show very good conversion rates [20]. It is, however, still important to be aware that incomplete bisulfite conversion may lead to false-positive results, especially when using very sensitive analytical methods, such as the widely used methylation-specific PCR (MSP) [21]. Likewise, one should be hesitant to use bisulfite-based methods when studying cells or

loci, which may have significant levels of 5-hydroxymethylcytosine as the bisulfite conversion cannot distinguish 5-methylcytosine and its oxidized derivatives [22].

Although a wide plethora of PCR-based methods are available for the analysis of DNA methylation, the vast majority do not provide information on the methylation status of individual alleles of genes [18]. Traditionally, bisulfite sequencing of individual clones has been used to obtain allelic methylation information, provided that a heterozygous SNP is found within the amplicon. This method is, however, labour-intensive, expensive, and may not be suitable for use in clinical settings.

We present here a protocol for allele-specific DNA methylation detection based on Pyrosequencing of methylation-specific PCR (MSP) products referred to as allelic MSP-Pyrosequencing. The methylation-specific primers are designed to surround a SNP used to distinguish amplification from the two alleles. Preferably, additional CpG sites as well as non-CpG cytosines should also be found in between the primers as a control for the amplification of fully methylated and bisulfite-converted molecules, respectively, to avoid false-positive results. When initially optimizing an allelic MSP-Pyrosequencing assay, we recommend running standards of known allelic methylation status to validate the ability of the assay to distinguish the methylation status of individual alleles. Compared to bisulfite sequencing of single clones allelic MSP-Pyrosequencing is fast, cost-effective, and does not suffer from PCR bias or cloning bias. While bisulfite sequencing of single clones takes several days to perform, allelic MSP-Pyrosequencing can be performed in less than 4 h. We have previously shown that allelic methylation patterns can be reliably analyzed in samples methylated at levels below 1 % when using 25 ng input DNA [23]. It would require sequencing of hundreds of clones to obtain a similar analytical sensitivity using bisulfite sequencing.

2 Materials

2.1 Preparation of Samples and Controls

1. Thermocycler (e.g., Veriti® 96-Well Fast Thermal Cycler, Life Technologies).
2. CpG Methyltransferase (M.SssI), reaction buffer, and S-adenosylmethionine (SAM) (e.g., New England Biolabs).
3. CpGenome™ Universal Methylated DNA (Millipore) (optional *see* Subheading 3.2).
4. CpGenome™ human non-methylated DNA (Millipore) (optional *see* Subheading 3.2).
5. Human genomic DNA from human blood (buffy coat) (Roche) (optional *see* Subheading 3.2).
6. Whole genome amplification kit (GE Healthcare) (optional *see* Subheading 3.2).

2.2 DNA Extraction

1. QIAamp DNA Mini/Micro Kit (Qiagen).
2. Minispin.
3. Thermomixer.
4. Ethanol (96–100 %).
5. Spectrophotometer, Bioanalyzer (Agilent Technologies) or equivalent.

2.3 Sodium Bisulfite Conversion

1. EZ DNA methylation kit (Zymo Research).
2. Minispin.
3. Thermocycler (e.g., Veriti® 96-Well Fast Thermal Cycler, Life Technologies).
4. Ethanol (96–100 %).

2.4 PCR Amplification

1. Thermocycler (e.g., Veriti® 96-Well Fast Thermal Cycler, Life Technologies).
2. PCR primers (purification by desalting).
3. Biotinylated PCR primer (purification by desalting).
4. PyroMark® PCR kit (Qiagen) (optional *see* Subheading 3.6).
5. LightCycler 480 High Resolution Melting Master (Roche) (optional *see* Subheading 3.6).

2.5 Gel Electrophoresis

1. Gel electrophoresis device.
2. UV transilluminator or imaging system.
3. Microwave oven.
4. TAE buffer: 40 mM Tris, 40 mM acetate, and 1 mM EDTA.
5. Agarose (molecular biology grade).
6. Ethidium Bromide.
7. DNA Ladder, range 50–500 bp.

2.6 Real-Time PCR Amplification and Melting Analysis

1. LightCycler® 480 instrument (Roche).
2. LightCycler® 480 Multiwell Plate 96 (Roche).
3. PCR primers (purification by desalting).
4. Biotinylated PCR primer (purification by desalting).
5. LightCycler® 480 SYBR Green I Master (Roche).

2.7 Pyrosequencing

1. PyroMark Q24 Instrument (Qiagen).
2. PyroMark Q24 Vacuum Workstation (Qiagen).
3. Millex-FG50 Filter Unit (Millipore).
4. PyroMark Q24 Vacuum Prep Troughs (Qiagen).
5. Thermomixer.
6. Heating device, for example, heating plate or thermoblock.

7. Thermoplate for sample preparation (Qiagen).
8. 96-Well PCR Plate, nonskirted, and strips or sealing foil.
9. Streptavidin Sepharose HP beads (GE Healthcare).
10. Pyrosequencing primer (purification by desalting).
11. Wash-, denaturation-, binding-, and annealing buffers (Qiagen).
12. PyroMark Gold Q24 Reagents (Qiagen).
13. PyroMark Q24 Cartridge (Qiagen).
14. PyroMark Q24 Plate (Qiagen).

3 Methods

The protocol presented here for allele-specific DNA methylation detection requires that the sample to be analyzed is heterozygous for a SNP found in the vicinity of the CpG sites to be analyzed. The SNP genotyping is not described within this chapter, but can be performed using Pyrosequencing as described [24], or by other standard genotyping methods. Homozygous samples can still be analyzed along with the heterozygous samples, but the methylation information acquired for these samples will not be allele-specific. A list of the most polymorphic SNPs found in CpG islands of a number of genes, for which methylation has been associated with various human malignancies can be found in Table 1. However, many more cancer relevant genes overlap with CpG islands, which contain common SNPs (Fig. 1).

The first step of the protocol is a PCR amplification of the region containing the CpG sites of interest as well as the SNP using biotinylated methylation-specific PCR primers. This allows Pyrosequencing to be used to quantify the methylation level of each of the two alleles. Simultaneously, information on the methylation status of additional CpG sites in between the primers as well as the bisulfite conversion status of non-CpG sites is provided and used as an internal quality control. The PCR amplification can be carried out as a standard PCR or as a real-time PCR to avoid gel electrophoresis and/or to gain quantitative methylation information. Because the primers are specific for the amplification of methylated DNA it is recommended only to perform Pyrosequencing of positive samples as determined by melting analysis or gel electrophoresis. The protocol is outlined in Fig. 2.

3.1 Assay Design

The assay design is a critical step for all PCR-based methods for DNA methylation analysis. In particular, when the assay is based on MSP primers, it is important to pay great attention to primer design to avoid amplification of unmethylated DNA or incompletely bisulfite converted molecules, which may lead to false-positive results. Primer design considerations for PCR-based methylation analysis

Table 1
Polymorphic SNPs found in CpG islands of cancer relevant genes

Gene symbol	Cancer type	SNP	Type	Average heterozygosity	Genomic position
<i>ADAMTS1</i>	Colorectal, pancreatic, prostate, NSCLC	rs445784	A/C	0.454 ± 0.145	chr21:28216692
		rs434857	G/T	0.448 ± 0.153	chr21:28217178
		rs402007	C/G	0.447 ± 0.154	chr21:28217320
		rs416905	A/G	0.447 ± 0.154	chr21:28217446
<i>APC</i>	Colorectal, bladder, breast, prostate, etc.	rs79896135	C/G/T	0.265 ± 0.250	chr5:112043282
<i>ATM</i>	Colorectal, breast	rs189037	A/G	0.500 ± 0.015	chr11:108093833
<i>BAP1</i>	RCC	rs15038	A/T	0.428 ± 0.176	chr3:52445031
<i>BRCA1</i>	Breast, ovarian, NSCLC	rs799906	C/T	0.495 ± 0.052	chr17:41278116
		rs8176071	-/ACA	0.447 ± 0.154	chr17:41278006-05
		rs799905	C/G	0.500 ± 0.014	chr17:41277187
<i>CCND2</i>	Breast, NSCLC	rs1049606	C/T	0.473 ± 0.113	chr12:4383036
<i>CDKN2A</i>	Breast, colorectal, leukemia, MDS, NSCLC, etc.	rs3814960	A/C/G/T	0.500 ± 0.000	chr9:21975017
<i>CDH1</i>	Breast, colorectal, leukemia, MDS, NSCLC, etc.	rs3743675	C/G	0.355 ± 0.227	chr16:68771547
		rs12932357	A/C	0.337 ± 0.234	chr16:68771905
<i>CDH13</i>	Breast, colorectal, leukemia, NSCLC, etc.	rs201022587	-/GCGT	0.392 ± 0.206	chr16:82660937-36
		rs200436545	-/GC	0.374 ± 0.217	chr16:82660945-44
		rs62040565	C/T	0.353 ± 0.228	chr16:82660946
		rs8060301	A/T	0.469 ± 0.120	chr16:82661744
<i>DAPK1</i>	Breast, colorectal, DLBCL, NSCLC, etc.	rs13300553	A/G	0.498 ± 0.029	chr9:90113273
<i>ELMO3</i>	NSCLC	rs12923138	A/C	0.495 ± 0.051	chr16:67233266
<i>GSTP1</i>	Bladder, breast, liver, prostate, etc.	rs145589007	-/G	0.500 ± 0.014	chr11:67351530-29
		rs1079719	A/G	0.434 ± 0.170	chr11:67351531
		rs8191443	-/G	0.499 ± 0.025	chr11:67351537-36
		rs4147581	C/G	0.491 ± 0.068	chr11:67351585
		rs2370143	C/T	0.384 ± 0.211	chr11:67351774
<i>HIC1</i>	Breast, colorectal, MDS	rs9901806	C/T	0.496 ± 0.044	chr17:1959822
<i>hMLH1</i>	Breast, colorectal	rs1800734	A/G	0.431 ± 0.172	chr3:37034946
<i>MGMT</i>	Breast, colorectal, glioblastoma, NSCLC, etc.	rs181536588	A/T	0.180 ± 0.240	chr10:131265388
<i>miR-124-1</i>	Breast, gastric, etc.	rs531564	C/G	0.245 ± 0.250	chr8:9760699

(continued)

Table 1
(continued)

Gene symbol	Cancer type	SNP	Type	Average heterozygosity	Genomic position
<i>MSH2</i>	Colorectal, RCC	rs2303426	C/G	0.473 ± 0.113	chr2:47630550
<i>PTEN</i>	Breast, glioblastoma	rs12573787	A/G	0.341 ± 0.233	chr10:89623716
<i>RASSF1</i>	Breast, NSCLC, prostate	rs4688725	A/C	0.394 ± 0.205	chr3:50378176
<i>RBI</i>	Glioblastoma, lymphoma, retinoblastoma, etc.	rs2252544	C/T	0.494 ± 0.053	chr13:48878292
<i>RUNX1</i>	Ovarian	rs743289	A/C	0.320 ± 0.240	chr21:36260588
		rs7281361	A/G	0.477 ± 0.106	chr21:36262427
<i>RUNX3</i>	Bladder, breast, colorectal, liver, NSCLC, etc.	rs71514255	A/G	0.498 ± 0.028	chr1:25256444
<i>SFRP1</i>	Bladder, breast, liver, prostate, NSCLC	rs3832595	–/CTG/GCT	0.434 ± 0.169	chr8:41166638-40
<i>SFRP2</i>	Breast, colorectal, gastric, liver, prostate	rs4145339	C/T	0.313 ± 0.242	chr4:154710481
<i>SOCS1</i>	Breast, colorectal, liver, MDS, etc.	rs243332	A/G	0.369 ± 0.220	chr16:11349678
		rs33977706	G/T	0.343 ± 0.232	chr16:11350155
<i>TWIST1</i>	Bladder, breast, colorectal	rs76446751	G/T	0.371 ± 0.219	chr7:19156443
		rs73079392	C/G	0.366 ± 0.221	chr7:19156597
<i>VHL</i>	DLBCL, pancreatic, RCC	rs779805	A/G	0.486 ± 0.082	chr3:10183337

Information was acquired from the UCSC genome browser and NCBI dbSNP build 138

NSCLC non-small cell lung cancer, MDS myelodysplastic syndrome, RCC renal cell carcinoma

are described in detail elsewhere [18]. In addition to general primer design considerations for MSP primers, the primers should be far enough apart to allow a sequencing primer to be designed within the amplicon; however, it may be beneficial to keep the amplicon below 200 bp (*see Note 1*). Also a SNP should be present in between the primers to allow allele-specific information to be obtained (*see Note 2*). The commercial Pyrosequencing design software has been developed for the design of methylation-independent PCR primers and we, therefore, recommend that the MSP primers are designed manually.

The sequencing primer can also easily be designed manually. The first step is to identify a suitable region within the amplicon where no SNPs are found, which will allow sequencing of the

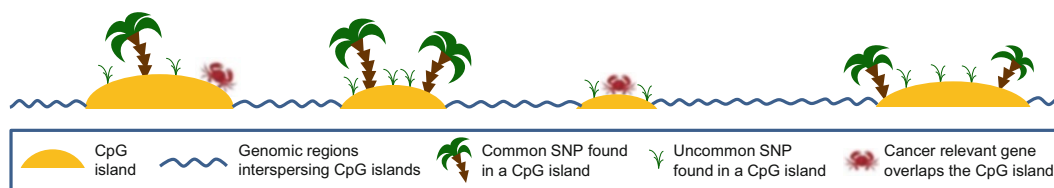


Fig. 1 Illustration of CpG islands in the human genome. Regions containing a relatively large proportion of CpG sites are normally referred to as CpG islands (defined by UCSC genome browser as regions of at least 200 bp with a GC content of at least 50 % and a ratio greater than 0.6 of the observed number of CpG sites to the expected number on the basis of the number of Gs and Cs in the segment). Some CpG islands contain common SNPs and overlap with a cancer relevant gene (the island to the left). The human genome contains 28,691 CpG islands (UCSC CGI file) and 12,160,626 common SNPs (NCBI dbSNP version 138 of SNPs with minor allele frequency ≥ 1 % filtered for nonbiallelic SNPs that has an exception comment or noncanonical chromosome name, e.g., chrUn). 22,494 CpG islands contain 93,539 unique common SNPs, 1,158 CGIs coincide with 607 unique cancer genes (human tumour suppressor genes, <http://bioinfo.mc.vanderbilt.edu/TSGene>), and 954 CGIs are overlapped by both a cancer gene (569 unique) and SNPs (4,273 unique)

region to be analyzed (*see Note 3*). Since only methylated DNA is amplified, CpG sites do not necessarily need to be avoided within the sequencing primer. However, it is important that the last four or five bases from the 3' terminus are unique in the amplicon, since as few as four consecutive nucleotides complementary to a sequence in the amplification product may increase the background signal and confound precise quantification [25]. Also the sequencing primer should be at least 15 bases and avoid long A or T stretches. The direction of the sequencing primer should be the opposite of the biotinylated MSP primer.

Another assay with primers specific for unmethylated DNA targeting the same CpG sites as the MSP primers should also be designed. These primers are used to confirm that a negative result is due to the sample being unmethylated and not because the DNA was too degraded to amplify. If using real-time PCR and quantitative data are desired, a control assay is needed to normalize for DNA input as the degradation of the samples during bisulfite treatment may vary. This assay should target a region depleted of CpG sites in a chromosomal region not likely to undergo copy number alterations in the samples to be studied, or target alu sequences depleted of CpG sites by evolutionary deamination [26, 27]. Finally, all of the designed primers should be verified for potential primer dimers and hairpin structures.

3.2 Preparation of Controls

Several positive and negative controls should be prepared for assay optimization procedures described in Subheading 3.3 as well as general use when performing allelic MSP-Pyrosequencing.

1. *Methylated DNA* can be purchased from a commercial company (e.g., CpGenome™ Universal Methylated DNA (Millipore), or a DNA sample (e.g., DNA extracted from peripheral blood

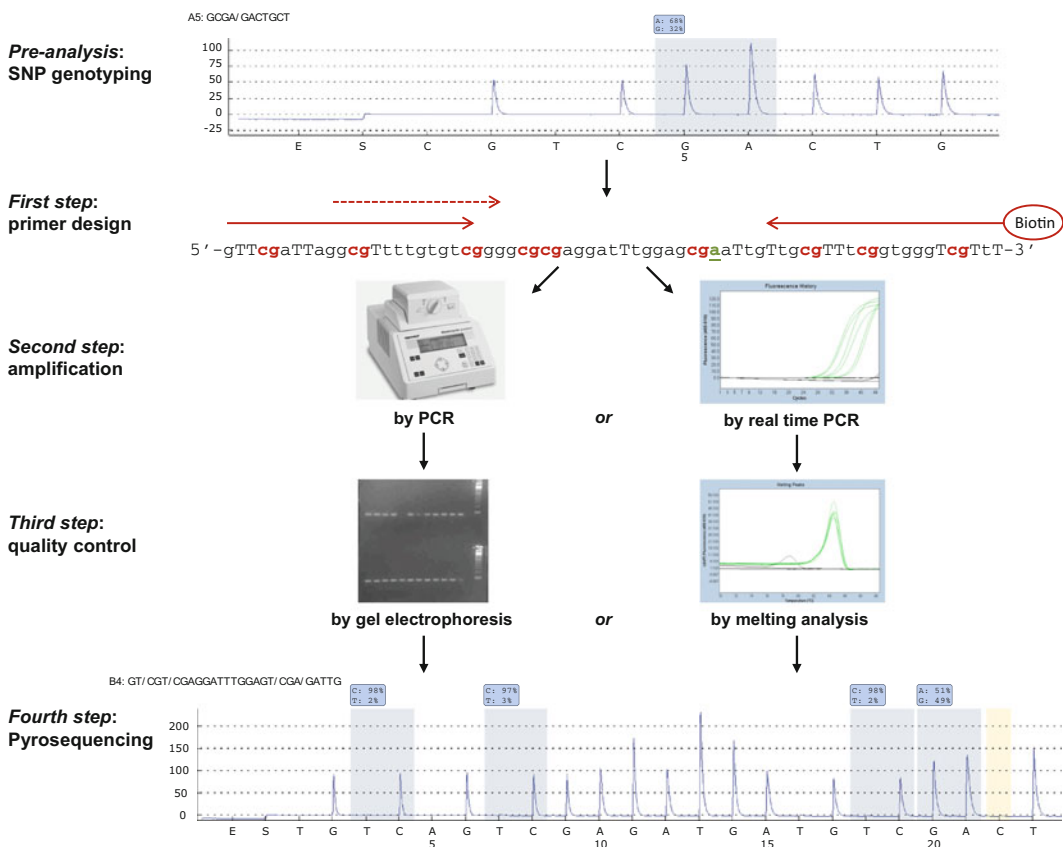


Fig. 2 Outline of the protocol. The first step of the protocol is to design methylation specific PCR primers (*solid arrows*) targeting CpG sites (*bold*) located within the CpG island of interest. The primers should surround the SNP (*underlined*), which is later used to distinguish the two alleles, and preferably additional CpG sites as well as non-CpG cytosines (*uppercase T*). One of the primers should be labelled with biotin, and the direction of this primer should be the opposite of the sequencing primer (*dashed arrow*). The next step is to amplify the bisulfite modified DNA from the samples to be studied. This can be done as a conventional PCR or as a real-time PCR followed by a quality control step using gel electrophoresis or melting analysis, respectively. For instance, melting analysis may be used to distinguish primer dimer formation from real amplification (small melting peak observed to the left of the peaks corresponding to the amplification product). Only positive samples should be subjected to the final step of the protocol, the Pyrosequencing. In the example shown here, it can be observed that amplification with the methylation specific primers occurred from both alleles (A:51 % and G:49 %). It can also be observed that the primers amplified methylated molecules only (methylation of the CpG sites was 98 %, 97 %, and 98 %, respectively). Finally, it can be observed that the bisulfite conversion was complete (no peak is observed at the non-CpG cytosine in the 3' end of the amplicon). It should be noted that allele-specific methylation is only obtained from samples being heterozygous for the SNP. Thus, the samples should be genotyped either before (as shown here) or after the methylation analysis

as described in Subheading 3.4) can be treated with a CpG Methyltransferase (M.SssI) purchased from a commercial company according to the manufacturer's instructions (*see Note 4*). The methylated DNA should be bisulfite treated as described in Subheading 3.5 before use.

2. *Unmethylated DNA* can also be purchased from a commercial company (e.g., Millipore or Qiagen). However, DNA extracted from the peripheral blood of a healthy blood donor is generally unmethylated for most CpG islands and if used we recommend a whole genome amplification of the DNA be performed as previously described [28] to ensure that the DNA is essentially unmethylated (*see Note 5*). The unmethylated DNA should be bisulfite treated as described in Subheading 3.5 before use.
3. *Unconverted DNA* can be any source of genomic DNA (e.g., DNA extracted from peripheral blood as described in Subheading 3.4). This genomic DNA sample should not be bisulfite treated.
4. *No Template Control*.
5. *Standards of known allelic methylation status* are prepared by treating samples of each possible genotype of the SNP used to distinguish the allelic methylation patterns with a CpG Methyltransferase (M.SssI) purchased from a commercial company, according to the manufacturer's instructions. DNA of each genotype can normally be obtained by screening a small panel of peripheral blood DNA samples. Finally, these three standards should be bisulfite treated as described in Subheading 3.5 before use.

3.3 Assay Optimization and Validation

We recommend that the MSP primers are initially purchased without biotin. Before purchasing a biotinylated primer, it should be verified that the primers produce a single strong and specific PCR product when amplifying bisulfite-converted methylated DNA, and do not amplify bisulfite-converted unmethylated DNA, unconverted DNA as well as the No Template Control (*see Note 6*). The PCR can be performed as a standard PCR followed by gel electrophoresis or as a real-time PCR followed by melting analysis as described in Subheadings 3.6 and 3.7. A serial dilution of bisulfite-converted methylated DNA into bisulfite-converted unmethylated DNA should also be assessed to determine the lower analytical sensitivity of the assay, and, if using real-time PCR, to determine the quantitative accuracy and PCR efficiency of the assay, if quantitative methylation data is desired.

Next, a biotinylated MSP primer is ordered, and a PCR is performed for the standards of known allelic methylation status, at least in triplicates (*see Note 7*). These samples are then subjected to Pyrosequencing as described in Subheading 3.8. This experiment is used to determine the ability of the assay to distinguish the different allelic methylation states. When plotting the results for standards of known allelic methylation status, three different clusters should be observed (Fig. 3).

If background levels above 10–15 % are observed at the SNP position for the standards having only one methylated allele, it may

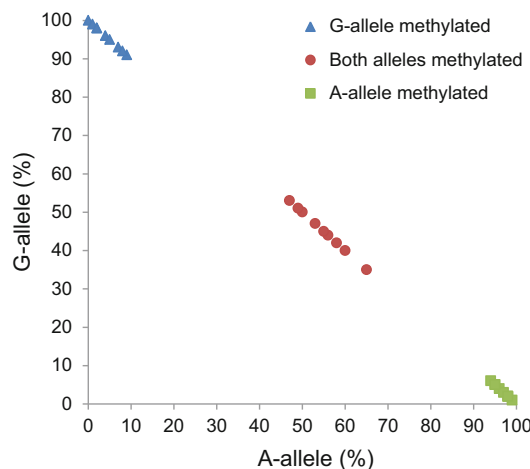


Fig. 3 Depending on the background levels in each individual assay the results for standards of known allelic methylation status should form three different clusters when plotting the quantitative Pyrosequencing results for each allele against each other. Not drawn to scale

be recommended to further optimize the assay or to redesign primers. Finally, it should be verified that the assay specific for unmethylated sequences, targeting the same CpG sites as the MSP primers, produce a single strong and specific PCR product when amplifying bisulfite-converted unmethylated DNA. If using real-time PCR and quantitative methylation data are desired, the quantitative accuracy and PCR efficiency of the control assay described in Subheading 3.1 should also be evaluated.

3.4 DNA Extraction

We normally use a commercial kit for DNA extraction from Qiagen (QIAamp DNA micro/mini kit) according to the manufacturer's protocol. Kits from other commercial companies or standard procedures such as phenol/chloroform extraction are also expected to perform well. The extracted DNA should be quantified using a spectrophotometer or other equivalent methods for the quantification of dsDNA, and the concentration should be adjusted to between 20 and 50 ng/ μ L depending on how much DNA to be used for sodium bisulfite conversion.

3.5 Sodium Bisulfite Conversion

For sodium bisulfite conversion we normally use the EZ DNA methylation kit (Zymo Research) or EpiTect bisulfite kit (Qiagen) according to the manufacturer's protocol. Kits from other suppliers are also expected to perform well. Depending on the number of analyses planned for each sample between 200 ng and 1 μ g DNA input should be used. The final elution volume should be within a range allowing a theoretical input of 20 ng bisulfite-converted DNA in the subsequent PCR amplification.

3.6 PCR Amplification

The PCR can be performed *either* as a standard PCR *or* as a real-time PCR if quantitative methylation data are needed and/or to avoid gel-electrophoresis.

1. When performing a standard PCR we normally use the PyroMark PCR kit from Qiagen. This kit contains a Master Mix with dNTPs, MgCl_2 , PCR buffer, and a HotStarTaq DNA polymerase, and also comes with an additive, the Q-solution, which may be helpful for sufficient amplification of very GC-rich regions, and a gel-tracking dye, the CoralLoad, which makes it easier to load the PCR product on a gel prior to Pyrosequencing analysis. For the reaction mixtures the Master Mix is used at a final concentration of 1 $\times$. Primer concentrations should be maximum 100–200 nM of each primer (*see Note 8*), and for each sample 20 ng bisulfite converted DNA is used. The final reaction volumes are 25 μL (*see Note 9*). The PCR cycling protocol starts with one cycle of 95 °C for 15 min, followed by 45 cycles of 95 °C for 20 s, annealing temperature determined in Subheading 3.3 for 20 s, 72 °C for 20 s, and one cycle of 72 °C for 10 min. After performing a standard PCR gel electrophoresis should be performed (jump to Subheading 3.7).
2. When performing a real-time PCR we normally use the LightCycler 480 High Resolution Melting Master from Roche. This kit contains a Master Mix with dNTPs, PCR buffer, FastStart Taq DNA Polymerase, and a dsDNA specific dye for product detection and characterization. MgCl_2 is supplied separately. For the reaction mixtures the Master Mix is used at a final concentration of 1 $\times$. Primer concentrations should be maximum 100–200 nM of each primer (*see Note 8*), and we normally use MgCl_2 at a final concentration of 2.5 mM. For each sample 20 ng bisulfite converted DNA is used. The final reaction volumes are 20 μL (*see Note 9*). The PCR cycling protocol starts with one cycle of 95 °C for 10 min, followed by 45 cycles of 95 °C for 20 s, annealing temperature determined in Subheading 3.3 for 20 s, 72 °C for 20 s. Fluorescence should be acquired to the SYBR Green I channel after each extension step. Then a 10 min hold at 72 °C, a 1 min hold at 95 °C, and a 1 min hold at 40 °C are completed before performing a melting analysis from 60 to 95 °C with 20 or more acquisitions per degree Celsius. Samples that did not amplify or produce a melting profile of the expected size (*see Note 10*) should not be subjected to Pyrosequencing. If quantitative data is desired, the samples should also be tested using the control assay described in Subheading 3.1.

If performing a real-time PCR jump to Subheading 3.8 afterwards. PCR products can be stored for several days at 4 °C or for several months at –20 °C.

3.7 Gel Electrophoresis

1. Prepare a 2 % agarose gel with ethidium bromide.
2. Load PCR products from Subheading 3.6, **step 1** and DNA size standard.
3. Run at 6 V/cm for 20–30 min depending on the size of the PCR product.
4. Place the gel on a UV Transilluminator or in an imaging system, and use the ladder to determine if the band is of the expected size.

Samples that did not produce a band of the expected size should not be subjected to Pyrosequencing. It is strongly recommended to separate pre- and post-PCR manipulations to reduce contamination risk.

3.8 Pyrosequencing

We have used the PyroMark Q24 Instrument and PyroMark Q24 Vacuum Workstation for which the workflow presented below is intended. However, allelic MSP-Pyrosequencing is expected to work equally well using the PyroMark Q96 ID (Qiagen). New users of the PyroMark Q24 instrument may benefit from reading the Pyromark Q24 user manual before continuing.

Before starting the workflow a Pyrosequencing run-file should be prepared using the PyroMark Q24 software.

1. To set up the assay select “new CpG assay,” paste in the sequence to analyze (*see* **Note 11**) and insert “/” between each of the possible bases at the variable positions (e.g., C/T before G at the CpG sites).
2. Press “generate dispensation order,” and click to insert bisulfite controls at the T bases derived from C.
3. Save the assay and select “new run.”
4. Now apply the assay to the number of wells that need to be analyzed and select an instrument method, which should be selected according to the reagents and cartridge that will be used for the run (*see* information on Qiagen’s webpage).
5. Print out the prerun information sheet and save the run file on a USB stick.
6. Transfer 10 μ L of the PCR product to a new PCR plate (*see* **Note 12**), and add 10 μ L of water to each sample.
7. Prepare a master mix containing per sample: 40 μ L binding buffer, 1 μ L Sepharose beads, and 19 μ L high-purity water. Mix well by flicking the tube or pipetting up and down, as it is crucial that the beads do not sediment. Then add 60 μ L of the master mix to each sample, and seal the plate with caps or sealing foil. Incubate in a thermomixer for 5–10 min at room temperature under constant mixing (2,000 rpm).

8. Meanwhile, switch on the heating device, and fill the PyroMark Q24 Vacuum Prep Troughs with 70 % ethanol, wash buffer, denaturation buffer, and high-purity water. Then prepare a master mix containing per sample: 0.75 μL sequencing primer (10 μM) and 24.75 μL annealing buffer. Mix by flicking the tube or pipetting up and down and add 25 μL to each of the wells, which are intended to be used, of a PyroMark Q24 Plate placed in the appropriate place within the workstation.
9. Switch on the vacuum pump, and apply vacuum to the tool by opening the vacuum switch. Then wash the filter probes by lowering the probes into the trough containing high-purity water next to the parking station.
10. Immediately after the samples have finished incubation, switch on the vacuum pump, and apply vacuum to the tool by opening the vacuum switch. Carefully lower the filter probes into the PCR plate to capture the beads containing immobilized template. Make sure no liquid remains in the PCR plate, and transfer the tool to the trough containing 70 % ethanol for 5–10 s.
11. Immediately hereafter transfer the tool to the trough containing denaturation solution for 5–10 s and finally to the trough containing wash buffer for 10–15 s.
12. Raise the tool to beyond 90° vertical for 5 s, to drain liquid from the filter probes, and switch off the vacuum pump and close the vacuum switch on the tool.
13. Release the beads in the plate containing the sequencing primer master mix, by shaking the tool gently from side to side for 30–60 s.
14. Transfer the tool to the trough containing high-purity water next to where the plate is positioned on the workstation.
15. Heat the PyroMark Q24 Plate containing the samples at 80 °C for 2 min using the heating device and the PyroMark Q24 Plate Holder (supplied with the workstation). Sealing of the plate is not necessary. Afterwards, allow the plate to cool down to room temperature for 4–5 min.
16. Dispense the nucleotides, substrate mix, and enzyme mix in the PyroMark Q24 Cartridge according to the handbook supplied with the reagents. The required volumes are found on the pre-run information sheet printed before starting the workflow.
17. Deposit the PyroMark Q24 Plate and the cartridge in the Pyrosequencer at their dedicated positions, and start the run by inserting the USB stick containing the run-file and selecting the run-file.

3.9 Data Analysis

The methylation status of the CpG sites and the bisulfite conversion is assessed using the PyroMark Q24 software in “CpG analysis mode.” For analysis of the SNP, switch to the “AQ analysis mode.” If the amplified methylated molecules are homogeneously methylated the CpG sites should be methylated at levels close to 100 %. However, if the amplified molecules are more heterogeneously methylated the levels will be lower. We have previously analysed the *DAPK1* gene promoter CpG island using allelic MSP-Pyrosequencing and found 8 out of 127 diffuse large B-cell lymphoma patients having an average methylation level between 70 and 85 %. These were interpreted as being heterogeneously methylated [23]. The potential for heterogeneously methylated molecules to be amplified is dependent on the annealing temperature used in the PCR [21]. Some CpG islands may be more prone to become heterogeneously methylated than others. Therefore, it may not be possible to define a universal cut-off to distinguish heterogeneous methylation from false-positives caused by the amplification of unmethylated molecules. If samples show failed bisulfite conversion this should be interpreted as missing data rather than the sample being unmethylated.

The SNP analysis is used to determine which one of the two alleles were amplified by the methylation-specific primers and, thus, which one of the alleles are methylated. If the data separate nicely into three different clusters as depicted in Fig. 3, it is straightforward to interpret the data. However, this is not likely to be the case in every situation when analysing biological samples. For instance, if analyzing DNA extracted from solid tumours, intratumour heterogeneity and aneuploidy with respect to the locus of interest may blur the picture and even cause the clusters to overlap. In that case a cut-off can be defined to distinguish samples being methylated at one allele and samples being methylated at both alleles. This cut-off could be defined as the mean background level in the standards having one allele methylated plus three standard deviations (i.e., if the mean background level is 4.2 % and the standard deviation is 3.1 % the cut-off should be at 13.5 %). Assuming that the background levels are normally distributed this cut-off will ensure that there is a less than 1 % chance of a sample being methylated at only one allele falls above the cut-off (*see Note 13*).

If using real-time PCR, it is possible to quantitatively estimate the fraction of methylated molecules within each sample analyzed; however, quantification strategies will not be described in detail here. We normally use relative quantification ($2^{-\Delta\Delta C_t}$) as previously described [28, 29]. The CpGenome™ Universal Methylated DNA (Millipore) can be used as calibrator (i.e., the sample for which the samples of unknown methylation levels are quantified relative to) [26, 28, 30]. For this purpose four different C_t -values are required; the C_t -values of the sample and the calibrator in the target gene assay, and the C_t -values of the sample and the calibrator in the control assay (*see Note 14*).

4 Notes

1. If the samples intended to be analyzed are of poor quality (e.g., DNA extracted from formalin-fixed paraffin-embedded (FFPE tissues)), it is advisable to keep the amplicon as short as possible and preferably below 100 bp.
2. The SNP should be as polymorphic as possible within the population to be analyzed to allow as many samples to be analyzed with allele-specificity as possible. SNP data can be obtained from NCBI's SNP database (<http://www.ncbi.nlm.nih.gov/SNP/>). In addition, it is important to keep in mind that C>T SNPs cannot be distinguished after bisulfite conversion. If a C>T SNP is intended to be used the MSP primers should be designed for the opposite strand where the SNP will be G>A. Likewise, if the SNP includes a C, which is part of a CpG site, the MSP primers should be designed for the opposite strand.
3. The region to be analyzed by Pyrosequencing should preferably include the SNP used to provide allele-specific information as well as CpG sites and non-CpG cytosines used as controls for the amplification of fully methylated and bisulfite converted template, respectively.
4. Treatment of genomic DNA with CpG Methyltransferase (M.SssI) enzymes may result in the methylation of cytosine outside of a CpG context at low levels (unpublished data). Methylation of non-CpG cytosines may interfere with primer binding and could be falsely interpreted as incomplete bisulfite conversion.
5. We have previously shown that a range of cancer-related genes may become methylated at low levels in the peripheral blood from healthy blood donors [21]. Performing a whole genome amplification of the DNA will create DNA, which is essentially unmethylated as the DNA polymerase does not distinguish between cytosines and methylated cytosines.
6. The PCR may have to be optimized in order to get a strong amplification signal from the positive control while not amplifying any of the negative controls. If amplification from unmethylated DNA or genomic DNA is observed the annealing temperature of the PCR should be increased. If amplification from no template control is observed the primer concentrations should be minimized. However, occasional primer dimer formation distinguishable by gel electrophoresis and melting analysis may be acceptable.
7. It is crucial to evaluate the quantitative accuracy, reproducibility, and background levels of the Pyrosequencing reaction at

the SNP position by analyzing standards of known allelic methylation status. If the background levels are low in the standards having only one methylated allele (e.g., below 5 %), and the standards having both alleles methylated are measured to be close to 50 % for both alleles, it may be sufficient to test these standards in triplicates. However, we recommend that this experiment is repeated at least once or twice to evaluate potential run to run variation, before testing valuable samples.

8. Higher primer concentrations may give rise to background signals in the Pyrosequencing reaction as excess biotinylated primer may reduce the capture of the amplification product on the streptavidin-coated beads.
9. A lower reaction volume can be used in order to spare reagents. However, at least 10 μ L should be used for the Pyrosequencing reaction, and when using standard PCR we use 4 μ L for gel electrophoresis.
10. The melting profile of the studied samples should preferably be similar to the profile of a methylated control sample. However, as allele-specific methylation patterns may differ and heterogeneous methylation may be present, it is likely that the melting profiles may deviate substantially. Therefore, we recommend that samples are subjected to Pyrosequencing when in doubt to prevent potential false-negative results.
11. The sequence to analyze starts immediately next to the 3' terminus of the sequencing primer and should preferably cover at least two CpGs, a non-CpG cytosine, and the SNP. There is no point in sequencing into the region covered by the amplification primer.
12. A PCR plate can be cut into four smaller plates of 3×8 wells. This is the correct format of the PyroMark Q24.
13. The background levels may depend on the quality of the samples analyzed. For instance, if the DNA is extracted from FFPE tissues higher background levels may be observed.
14. When using the $2^{-\Delta\Delta C_t}$ relative quantification approach the PCR efficiency of the target gene assay and the control assay should be approximately equal. If this is not the case, it may be beneficial to use an algorithm that incorporates PCR efficiency [31].

Acknowledgments

We acknowledge Helene Myrtue Nielsen for critical reading of this manuscript. This work was supported by a grant from the Danish Cancer Society.

References

1. Dobrovic A, Kristensen LS (2009) DNA methylation, epimutations and cancer predisposition. *Int J Biochem Cell Biol* 41:34–39
2. Reik W, Dean W, Walter J (2001) Epigenetic reprogramming in mammalian development. *Science* 293:1089–1093
3. Deaton AM, Bird A (2010) CpG islands and the regulation of transcription. *Genes Dev* 25:1010–1022
4. Ross JP, Rand KN, Molloy PL (2010) Hypomethylation of repeated DNA sequences in cancer. *Epigenomics* 2:245–269
5. Armstrong L (2012) Epigenetic control of embryonic stem cell differentiation. *Stem Cell Rev* 8:67–77
6. Smallwood SA, Kelsey G (2012) De novo DNA methylation: a germ cell perspective. *Trends Genet* 28:33–42
7. Chang SC, Tucker T, Thorogood NP et al (2006) Mechanisms of X-chromosome inactivation. *Front Biosci* 11:852–866
8. Kaneda M (2011) Genomic imprinting in mammals-epigenetic parental memories. *Differentiation* 82:51–56
9. Tycko B (2010) Allele-specific DNA methylation: beyond imprinting. *Hum Mol Genet* 19:R210–R220
10. Feng J, Fan G (2009) The role of DNA methylation in the central nervous system and neuropsychiatric disorders. *Int Rev Neurobiol* 89:67–84
11. Javierre BM, Hernando H, Ballestar E (2011) Environmental triggers and epigenetic deregulation in autoimmune disease. *Discov Med* 12:535–545
12. Kwok JB (2010) Role of epigenetics in Alzheimer's and Parkinson's disease. *Epigenomics* 2:671–682
13. Baylin SB, Jones PA (2011) A decade of exploring the cancer epigenome: biological and translational implications. *Nat Rev Cancer* 11:726–734
14. How Kit A, Nielsen HM, Tost J (2012) DNA methylation based biomarkers: practical considerations and applications. *Biochimie* 94:2314–2337
15. Fenaux P, Mufti GJ, Hellstrom-Lindberg E et al (2009) Efficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: a randomised, open-label, phase III study. *Lancet Oncol* 10:223–232
16. Treppendahl MB, Kristensen LS, Grønbæk K (2014) Predicting response to epigenetic therapy. *J Clin Invest* 124:47–55
17. Kristensen LS, Hansen LL (2009) PCR-based methods for detecting single-locus DNA methylation biomarkers in cancer diagnostics, prognostics, and response to treatment. *Clin Chem* 55:1471–1483
18. Kristensen LS, Treppendahl MB, Grønbæk K (2013) Analysis of epigenetic modifications of DNA in human cells. *Curr Protoc Hum Genet*. Chapter 20:Unit20 2
19. Tost J, Gut IG (2007) DNA methylation analysis by Pyrosequencing. *Nat Protoc* 2:2265–2275
20. Hernandez HG, Tse MY, Pang SC et al (2013) Optimizing methodologies for PCR-based DNA methylation analysis. *Biotechniques* 55:181–197
21. Kristensen LS, Raynor MP, Candiloro I et al (2012) Methylation profiling of normal individuals reveals mosaic promoter methylation of cancer-associated genes. *Oncotarget* 3:450–461
22. Huang Y, Pastor WA, Shen Y et al (2010) The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS One* 5:e8888
23. Kristensen LS, Treppendahl MB, Asmar F et al (2013) Investigation of MGMT and DAPK1 methylation patterns in diffuse large B-cell lymphoma using allelic MSP-pyrosequencing. *Sci Rep* 3:2789
24. Royo JL, Galan JJ (2009) Pyrosequencing for SNP genotyping. *Methods Mol Biol* 578:123–133
25. Dejeux E, El abdalaoui H, Gut IG et al (2009) Identification and quantification of differentially methylated loci by the pyrosequencing technology. *Methods Mol Biol* 507:189–205
26. Kristensen LS, Nielsen HM, Hager H et al (2011) Methylation of MGMT in malignant pleural mesothelioma occurs in a subset of patients and is associated with the T allele of the rs16906252 MGMT promoter SNP. *Lung Cancer* 71:130–136
27. Weisenberger DJ, Campan M, Long TI et al (2005) Analysis of repetitive element DNA methylation by MethyLight. *Nucleic Acids Res* 33:6823–6836

28. Kristensen LS, Mikeska T, Krypuy M et al (2008) Sensitive melting analysis after real time-methylation specific PCR (SMART-MSP): high-throughput and probe-free quantitative DNA methylation detection. *Nucleic Acids Res* 36:e42
29. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C(T)}$ method. *Methods* 25:402–408
30. Kristensen LS, Wojdacz TK, Thestrup BB et al (2009) Quality assessment of DNA derived from up to 30 years old formalin fixed paraffin embedded (FFPE) tissue for PCR-based methylation analysis using SMART-MSP and MS-HRM. *BMC Cancer* 9:453
31. Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 29:e45

Free ebooks ==> www.Ebook777.com

Chapter 21

SNP-Based Quantification of Allele-Specific DNA Methylation Patterns by Pyrosequencing®

Florence Busato and Jörg Tost

Abstract

The analysis of allele-specific DNA methylation patterns has recently attracted much interest as loci of allele-specific DNA methylation overlap with known risk loci for complex diseases and the analysis might contribute to the fine-mapping and interpretation of non-coding genetic variants associated with complex diseases and improve the understanding between genotype and phenotype. In the presented protocol, we present a method for the analysis of DNA methylation patterns on both alleles separately using heterozygous Single Nucleotide Polymorphisms (SNPs) as anchor for allele-specific PCR amplification followed by analysis of the allele-specific DNA methylation patterns by Pyrosequencing®. Pyrosequencing is an easy-to-handle, quantitative real-time sequencing method that is frequently used for genotyping as well as for the analysis of DNA methylation patterns. The protocol consists of three major steps: (1) identification of individuals heterozygous for a SNP in a region of interest using Pyrosequencing; (2) analysis of the DNA methylation patterns surrounding the SNP on bisulfite-treated DNA to identify regions of potential allele-specific DNA methylation; and (3) the analysis of the DNA methylation patterns associated with each of the two alleles, which are individually amplified using allele-specific PCR. The enrichment of the targeted allele is re-enforced by modification of the allele-specific primers at the allele-discriminating base with Locked Nucleic Acids (LNA). For the proof-of-principle of the developed approach, we provide assay details for three imprinted genes (*IGF2*, *IGF2R*, and *PEG3*) within this chapter. The mean of the DNA methylation patterns derived from the individual alleles corresponds well to the overall DNA methylation patterns and the developed approach proved more reliable compared to other protocols for allele-specific DNA methylation analysis.

Key words Pyrosequencing®, Allele-specific PCR, Bisulfite, SNP, locked nucleic acid

1 Introduction

It has been assumed that with the exception of a small number of imprinted genes as well as genes on the X chromosome in females, most genes are expressed equally from both parental alleles. Similarly, DNA methylation, a key mechanism of gene regulation [1], was believed to be equally present on both alleles. However, allele-specific DNA methylation patterns including both sequence-dependent profiles associating with genetic variation in *cis* (and less frequently in *trans*) as well as stochastic allelic

methylation profiles have been realized to be a widespread phenomenon [2, 3]. Changes in DNA methylation patterns between the two alleles are often subtle and heterogeneous and might be—as DNA methylation profiles in general—specific for a cell type and/or a developmental stage. Nonetheless, the analysis of allele-specific DNA methylation patterns has recently attracted much interest as loci of allele-specific DNA methylation do seem to overlap with known at risk loci for complex diseases and might contribute to the fine-mapping and interpretation of noncoding genetic variants associated with complex diseases and improve the understanding between genotype and phenotype [4].

Pyrosequencing® [5, 6] is a quantitative real-time sequencing method that is frequently used for genotyping [7] as well as for the analysis of DNA methylation patterns [8–10] (*see* also Chapters 13–16). Pyrosequencing is based on the presence or absence of the incorporation of a nucleotide during primer extension [5, 11]. In contrast to Sanger sequencing, which relies on the random incorporation of fluorescent ddNTPs during primer extension steps, only one specific nucleotide is present at any time in the Pyrosequencing reaction. As described in detail in Chapter 1, the pyrophosphate (PPi) released following nucleotide incorporation is used as substrate in combination with adenosine 5' phosphosulfate (APS) by an ATP sulfurylase to produce ATP [12]. The latter is in turn used by luciferase to oxidize luciferin into oxyluciferin resulting in a light emission proportional to the amount of incorporated nucleotide [12]. The limit of detection of Pyrosequencing has been evaluated around 5 % for the minor allele, which is thus far more sensitive than Sanger sequencing [13].

In the protocol described in this chapter, we present a method to analyze the DNA methylation patterns on both alleles separately using heterozygous Single Nucleotide Polymorphisms (SNPs) as anchor for allele-specific PCR amplification. The enrichment of the targeted allele is re-enforced by modification of the allele-specific primers at the allele-discriminating base with Locked Nucleic Acids (LNA). LNA are nucleosides similar to nucleic acids, but they are “locked” with a 2'-O-4'-C methylene bridge in the ribose ring. This bridge increases the monomer's thermal stability, reduces its flexibility, and thereby increases the hybridization interactions of the base by maintaining the ideal conformation for binding. The resulting duplex is also more stable. LNA can contain the traditional bases found in DNA and RNA and has been used in many applications [14]. LNAs have also proven to be powerful tools for the selective amplification of subpopulations of molecules including the amplification of mutated molecules in a large background of wild-type molecules [15] (*see* also Chapter 7).

The presented protocol is applicable to the analysis of any gene containing potentially allele-specific DNA methylation patterns. For the proof-of-principle, we provide assay details for three

imprinted genes (*IGF2*, *IGF2R*, and *PEG3*) within this chapter. Imprinting concerns a small number of genes (100–200), which are characterized by a parent-of-origin dependent expression of the corresponding gene often associated with allele-specific DNA methylation in so called imprinting control regions, and plays an essential role in development, pre- and postnatal life [16]. Although representing only a fraction of allele-specific methylation patterns, it constitutes a paradigm and might be useful for setting-up the protocol in the laboratory.

Alternative protocols for allele-specific methylation analysis by Pyrosequencing have previously been devised. However, allele-discrimination was performed using a standard PCR followed by the use of an allele-specific Pyrosequencing primer [17]. In our hands, the allele-discrimination did work better if performed at the stage of the PCR amplification and the mean methylation values of the two alleles correlated better with the overall DNA methylation profile obtained by standard Pyrosequencing. This might be due to the fact that because of the thermal instability of the enzymes used in the Pyrosequencing reaction, this reaction needs to be performed at low temperature (i.e., 28 °C), which might negatively influence the discrimination between the two alleles. Furthermore, our method allows assessing the spatial extent of the allele-specific DNA methylation. In contrast to the above-described approach, which restricts the allele-specific analysis to the read-length of a Pyrosequencing run, i.e., in most cases less than 100 bases, the generation of an allele-specific amplification product allows the use of multiple Pyrosequencing primers and thereby permits to cover the entire amplification product. In addition, the discrimination at the PCR stage allows for more options of the assay design as the allele-specific primer can be placed at 3' or 5' end of the amplification product and amplification can be performed on either the forward or the reverse strand which are no longer complementary after bisulfite treatment (yielding thus altogether four possibilities for primer selection).

An alternative approach to assess the correlation between DNA methylation and genetic variation is based on the amplification of molecules with a specific DNA methylation patterns (usually completely methylated molecules) and analyzing quantitatively the proportions of the associated alleles by measuring the ratio of the two alleles of a heterozygous SNP by Pyrosequencing after this methylation-specific PCR [18] (*see* also Chapter 20).

2 Materials

2.1 Assay Design

Design of PCR primers for amplification of genomic DNA: Pyrosequencing Assay Design software (Version 1.0.6, Biotage, now Qiagen).

- Design of PCR primers for the amplification of bisulfite-treated DNA: MethPrimer [3] (<http://www.ucsf.edu/urogene/methprimer/index1.html>).
- Design of Pyrosequencing primers: Pyrosequencing Assay Design software (Version 1.0.6, Biotage, now Qiagen) or manually.
- Design of PCR primers using Primer 3 [19, 20] (<http://bioinfo.ut.ee/primer3-0.4.0/>).

2.2 Bisulfite Treatment

1. EpiTect 96 Bisulfite Kit (Qiagen).
2. Ethanol.
3. 96-Well Microtiter Plates with conical wells (e.g., Abgene AB800).
4. 2 mL microtubes.
5. Pipettes 25 and 50 mL.
6. PVC Reservoir.
7. Thermocycler.
8. Centrifuge for microtiter plates.

2.3 Quantification of Bisulfite-Treated DNA

1. Commercial fully unmethylated or methylated DNA (e.g., Qiagen).
2. KAPA PROBE Fast qPCR Master Mix (2×) (CliniSciences).
3. Probe for real-time PCR amplification 5'-6FAM- CCT ACC TTA+ AC+C T+C+C C- BBQ, modified from [19]; (BBQ=Black Berry Quencher, +=LNA-modified nucleotide).
4. Forward (5'-GGT TAG GTA TAG TGG TTT ATA TTT GTA ATT TTA GTA) and reverse (5'-ATT AAC TAA ACT AAT CTT AAA CTC CTA ACC TCA) primer for real-time PCR amplification [19].
5. 96-Well real-time thermocycler with associated software (e.g., LightCycler480, Roche Molecular Diagnostics).
6. Plates for real-time thermocycler.

2.4 PCR Amplification

1. Reference samples for all genotypes of a SNP under investigation.
2. Primers for PCR amplification (unmodified, containing LNAs, and/or biotinylated).
3. Polymerase for amplification of bisulfite-treated DNA (e.g., HotStar Taq DNA polymerase, Qiagen).
4. Polymerase for amplification of genomic DNA (e.g., Platinum Taq DNA Polymerase High Fidelity; Thermo Scientific).
5. dNTPs (8 mM).
6. Gradient Thermocycler.

All reagents should be stored at -20 °C.

2.5 Sample Preparation

1. Streptavidin Sepharose HP beads.
2. Vacuum preparation tool (Qiagen) using the corresponding filter probes and Prep Troughs.
3. Binding buffer: 10 mM Tris-Cl, 2 M NaCl, 1 mM EDTA, 0.1 % Tween 20; pH 7.6.
4. Ethanol 70 %.
5. Denaturing solution: 0.2 M NaOH.
6. Washing buffer: 10 mM Tris-acetate; pH 7.6.
7. Annealing buffer: 20 mM Tris-acetate, 2 mM Mg-Acetate; pH 7.6.
8. 96-Well Microtiter Plates with conical wells (e.g., AbGene AB800).
9. Sealing tape.
10. Thermomixer or similar (room temperature).
11. Heating device, for example, heating plate or thermoblock.
12. PSQ low microtiter plate for sample preparation (Qiagen).
13. Primers for Pyrosequencing.
14. All reagents used for this step should be stored at room temperature except for the streptavidin-coated Sepharose beads (+4 °C) and the Pyrosequencing primer (−20 °C).

2.6 Pyrosequencing Reaction

1. Pyrosequencer PyroMark® Q96 MD (Qiagen).
2. Cartridge for reagent dispensation: PyroMark Q96 HS 96 Dispensing Tip Holder (Qiagen).
3. PyroMark Q96 HS Nucleotide Tip (Qiagen).
4. PyroMark Q96 HS Q96 Reagent Tip (Qiagen).
5. Pyrosequencing Kit for genotyping: PyroMark GOLD Q96 Reagents (Qiagen).
6. Pyrosequencing Kit for DNA methylation analysis: PyroMark GOLD Q96 SQA Reagents (Qiagen).
7. PyroMark Q96 ID Software 2.5 (Qiagen).
8. Q-CpG software (Qiagen).

3 Methods

3.1 Assay Design

The assay design is the most critical step for DNA methylation analysis by Pyrosequencing in general (*see* also Chapter 13). The procedure for the design of the Pyrosequencing assay for the genotyping on genomic DNA is described in Subheading 3.1.1, while the design on bisulfite-treated DNA using MethPrimer [21] as performed in our laboratory is described in Subheading 3.1.2. Primers are positioned to amplify the target region independently of its methylation status.

3.1.1 *Design of a Genotyping Assay*

1. Identify a target SNP in the region of interest that is highly polymorphic in the population of interest using public databases or in-house available genotyping data.
2. Search for the SNP of interest using the UCSC (University of California Santa Cruz) Genome Browser or any other suitable resource and download the sequence of the region of interest of approximately 400 base pairs around the SNP of interest.
3. Verify that the sequence does not contain any other SNP in the population of interest.
4. Design the PCR primers using Pyrosequencing Assay Design software or the freely available Primer 3 software. The length of the amplification product should be around 100–150 bp.
5. Report in brackets or mark the SNP of interest in order to include the target region in the proposed assays. Primer size should be around 20 nucleotides and the T_m around 60 °C.
6. Investigate the presence of potential polymorphic positions such as SNPs underlying the annealing sites as well as primer specificity using a BLAT search on the UCSC Genome Bioinformatics portal. Discard the assays if amplification primers anneal to potential polymorphic sites and/or when both primers are highly complementary to several regions of the genome.
7. Design the Pyrosequencing primers using the commercial PSQ assay design software.
8. The sequencing primer should not anneal to any polymorphic and/or mutation site and should contain at least two nucleotides not overlapping with the PCR primers on its 3' end to avoid any unspecific signals or annealing to potential primer dimers.
9. At least the last four bases of the 3' end of the sequencing primers should be unique in the amplification product to avoid the generation of background signals, which could interfere with accurate quantification.
10. If possible, select the orientation of the sequencing reaction to avoid the presence of “T” mutations, as the complementary Pyrosequencing reaction will incorporate an “A” nucleotide. Adenosinetriphosphates can be used as substrate by the luciferase despite the use of dATP α S and can induce a high background signal [22].
11. Select the orientation of the sequencing reaction according to the ease of interpretation of the Pyrograms in one direction or the other around the SNP and include a biotin on the 5' end of one primer, depending on the direction of the Pyrosequencing assay.

12. Do not place the Pyrosequencing primer too far away from the SNP as the PyroMark Gold Q96 Reagents used for genotyping experiments can sequence only 30–40 nucleotides.

*3.1.2 Design
of the Pyrosequencing
Assay for DNA Methylation
Analysis*

The assay design is probably the most critical step for Pyrosequencing-based DNA methylation analysis. Great attention should therefore be paid to the experimental design, as this will crucially influence the successful outcome of the assay. Many standard software tools developed for conventional PCR cannot handle primer design on bisulfite-converted DNA due to the lower complexity of bisulfite-treated DNA. However, some commercial and freely available software have been designed specifically for this purpose such as MethPrimer [21], Bisearch [23], or MethylPrimer Express (Life Technologies).

1. Enter the sequence in MethPrimer and select the SNP as target to ensure it is included in the amplification product (*see Note 1*).
2. Primers should be approximately 30 nucleotides in length and the optimal size of an amplification product is ~250 bp, although we successfully analyzed DNA methylation in PCR products up to 350 bp in length.
3. Place primers in a region containing four or more cytosines that have been converted during bisulfite treatment to ensure that they are only complementary to completely converted DNA as the chemical treatment is rarely 100 % complete.
4. Primers should preferentially contain no CpG positions. If this cannot be avoided the maximum number of CpG positions covered should be restricted to one, and this CpG position should not be included in the last five bases from the 3' terminus to avoid preferential amplification.
5. To ensure specificity, palindromes within primers and complementary sequences between primers as well as degenerated bases or the use of inosine bases should be avoided.
6. The generated amplification product should be verified for the presence of potential polymorphic positions such as SNPs underlying the annealing sites for the amplification primers. We strongly recommend the re-design of amplification primers annealing to potentially polymorphic sites (even at the 5' terminus).
7. Identify sequence stretches in the amplification product where a Pyrosequencing primer can be positioned and at least the last five bases from the 3' terminus do not overlap with any other potentially variable position including CpGs and SNPs that are retained after bisulfite treatment (*see Note 2*).

8. Verify the last four or five bases from the 3' terminus to be unique in the amplification product by using, for example, the MS Word "Find" tool. As few as four consecutive nucleotides complementary to a non-targeted stretch of sequence in the amplification product might add to background signal confounding precise quantification.
9. Check successfully designed primers for primer dimers and possible hairpin structures.
10. As read-lengths of up to 120 bases can be achieved with the PyroGold SQA kit, primers can be positioned in nonpolymorphic regions next to the variable region using part of the nucleotide dispensations to approach the region of interest.
11. At least one cytosine not followed by a guanine should be included in the dispensation order to control for complete bisulfite conversion.
12. The direction of the Pyrosequencing primer defines which of the amplification primers needs to be biotinylated. This primer should be checked carefully not to form any hairpin structure.
13. If several Pyrosequencing primers are required to cover the region of interest, analysis of a few CpG positions by more than one Pyrosequencing primer improves confidence into the acquired results and helps to detect potential technical artifacts.

3.1.3 Design of the Pyrosequencing Assay for Allele-Specific DNA Methylation Analysis

As Methprimer does not permit the selection of specific primers, a slightly modified strategy has to be adapted.

1. Create a bisulfite-treated sequence either manually or by pasting the sequencing into MethPrimer.
2. Paste the sequence into Primer 3 and depending on the sequence context choose a forward or reverse amplification primer terminating with its last base on the SNP.
3. Repeat the procedure with primers of different length and orientation and evaluate the obtained primer pairs for the best combination.
4. Order the two primers complementary to each of two alleles and modify the last base from the 3' end of the primer (corresponding to the targeted SNP) into a LNA (*see Note 3*).
5. Design the Pyrosequencing primer as described in Subheading 3.1.2.

3.2 Identification of Heterozygous SNPs by Pyrosequencing

3.2.1 Amplification of Genomic DNA

In this step samples are genotyped to identify informative samples, i.e., samples heterozygous for one or a panel of SNPs.

1. Determine the optimal annealing temperature for the amplification of the target sequence designed in Subheading 3.1.1. This is performed on a gradient thermocycler using genomic DNA as template.

2. Typical PCR reaction conditions are 10–20 ng of genomic DNA, 1× HIFI Platinum Taq Buffer, 1 mM MgSO₄, 0.1 mM of each dNTP, 1.5 U of HIFI Platinum Taq Polymerase, and 200 nM of forward and reverse primers in a total volume of 25 µL.
3. Typical PCR cycling conditions include an initial denaturation step performed for 4 min at 95 °C, followed by 50 cycles of 30 s denaturation at 95 °C, 30 s of a gradient of annealing temperatures ranging from 50 to 70 °C, and 15 s elongation at 72 °C. Final extension is performed for 4 min at 72 °C.
4. Deposit 5 µL of the PCR product supplemented with 1× loading dye as well as a DNA size standard on a 2 % agarose gel including a fluorescent dye (example: ethidium bromide). Perform a horizontal electrophoresis in TBE 1× at 5 V/cm during 1 h. The specificity of the amplifications is assessed by the evaluation of the length of the PCR product and the highest temperature allowing a strong and specific amplification product is selected as the optimal annealing temperature.
5. Amplify the target sequence in the biological samples in a total volume of 25 µL containing 10–20 ng of genomic DNA, 1× HIFI Platinum Taq Buffer, 1 mM MgSO₄, 0.1 mM of each dNTP, 1.5 U of HIFI Platinum Taq Polymerase, and 200 nM of forward and reverse primer.
6. PCR cycling conditions include an initial denaturation step for 4 min at 95 °C, followed by 50 cycles of 30 s denaturation at 95 °C, 30 s at the optimal annealing temperature defined in **steps 2–4** (*see* Table 1 for examples), and 15 s elongation at 72 °C. Final extension is performed for 4 min at 72 °C.
7. Verify 5 µL of the PCR products on a 2 % agarose electrophoresis gel as described in **step 4**.
8. PCR products can be stored at +4 °C for several days or at –20 °C for several months.

3.2.2 Sample Preparation for Pyrosequencing

This step allows the preparation of the PCR products for Pyrosequencing experiments. It includes the purification and the denaturation of the amplification product, which will be rendered single stranded, and the hybridization of the sequencing primer (*see* **Note 4**).

1. Transfer 10 µL of PCR product into a standard skirted 96-well PCR plate and add 40 µL of binding buffer, 2 µL of Sepharose beads and 28 µL of water. Seal the plate and incubate for 10 min at room temperature under constant mixing (1,400 rpm) for the capture of the biotinylated PCR products by the Sepharose beads.

Table 1
Sequences and assay details for the analysis of *IGF2*, *IGF2R*, and *PEG3*

<i>IGF2</i> (rs680)	PCR primers	Amplified sequence position	Mg (mM)	Tp (°C)	Cycles	Pyrosequencing primer	PCR Vol (μL)	Amplified sequence
Genotyping PCR	5'-TGGACTTGAGTCCCT GAACC	chr11: 2153571– 2153671	0.5	56	50	5'-AGCAAGAGAA AAGAAGG	10	TGGAC TTGAGTCCCT GAACCAAGCAA AGAGAAAAGA
	5'-Biotin-TGAAAATTCCC GTGAGAAGG							AGGCCCCAG AAATCACAGG TGGGCAYGTC GCTCTACCG CCATCTCCCT TCTCACGGGA ATTTCA
Reference PCR	5'-GGTTTGGATTGAGTT TTTGAATTA	chr11: 2153454– 2153676	1.6	64	50	5'-TTAGAAATTATA GGTGGGTA	10	GGTTTGGATTGAGTTTTT GAATTAGTAAAGAGAAAAGAA
	5'-Biotin-ATCCTCCCTCC TTTAATCTTACTA							GGRTTTAGAAATTATAGGT GGTAYGYGTSTTATYG TTATTTTTTTTTTAYGGGAA TTTTTAGGGTAAATTGGT ATTYGAAATAGTAATAATT TAGATTGGTTTTTTTATTTTT TTTTTATTATTAATAAATTAT AGAGTAGTTAGAGGGATT AGTAAGATTAAAGGAGGGAGGAT
Allele-specific PCR	5'-GAATTAGTAAAGAG AAAAGAAG(G)	chr11: 2153407– 2153657	1.2	58,3	35	5'-TTAGAAATTAT AGGTGGGTA	10	GAATTAGTAAAGAGAAAAGAA GGRTTTAGAAATTATAGGT GGGTAYGTCGTSTTATCG
	ou 5'-GAATTAGTAAAGAG AAAAGAAG(A) 5'-Biotin-TCRTACCAATT ACATTTCAATTACA			58	15		15	TTATTTTTTTTTTAYGGGAAT TTTAGGGTAAATTGTTA TTYGAAAATAGTAATAATT TAGATTGGTTTTTTTATTTTT TTTTTATTATTAATAAATTAT AGAGTAGTTAGAGGGATT AGTAAGATTAAAGGAGGGGA GGATAGAGTATGAAAATTAA AATTTATGTAATGAAATGT AATTGGTACGA

PEG3 (rs3795005)	Amplified sequence position	Mg (mM)	Tp (°C)	Cycles	Pyrosequencing primer	PCR Vol (µL)	Amplified sequence
Genotyping PCR	5'-CACAGGCACTGG TTTGAATG 5'-Biotin-CAGCAAG ATGATGCCACAG	0.5	61	50	5'-GCAGAGGACT GAGTGA CTGA	10	CACAGGCACTGGTTGAATGACT GCTGGGTTGGCTGTAGGGCCGGGTG CAGAGGACTGAGTGA CTGAYGGGG TCTGCMTGCCCCCGCCACTCTG TGGCATCATCTTTGCTG
Reference PCR	5'-TTTGGTTTATTGT GGTGTAATTTT 5'-Biotin-ACCAACCCAA CTATCATTCAAAATA	1.6	61	50	5'-GTTGGGTT GGTTGTAG 5'-TGGTTGTAA TGGGGTA	10	TTTGGTTTATTGTGGTGTAAATTTTG GTTGTAATGGGCTAYGTGAGGGTGTAGT TATATAGGTA TTGGTTTGAATGATTG TTGGGTTGGTTGTAGGGTYGGGTGA GAGGATTGAGTGATTGAYGGGGTTTGT WTGGTTTYGTTTATTTTGTGGTATTAT TTTGTGTGTTATTAGGTGGTAGTTAY GGTGATTGTGTAGGTGTATTATTGAA TGATAGTTGGGTGGT
Allele-specific PCR	5'-GGTGTAATTTT GGTTGTAATGG 5'-Biotin-ATACCACAAAA TAAACRAAACCA(A) ou 5'-Biotin-ATACCACAAAA TAAACRAAACCA(T)	1.0	57,6	33	5'-GTTGGGTT GGTTGTAG 5'-TGGTTGTAA ATGGGGTA	20	GGTGAATTTTGGTGTGAATGG GGTAYGTGAGGGTGTAGTTATATA GGTATTGGTTTGAATGATTGTT GGGTTGGTTGTAGGGTYGGGT GTAGAGGATTGAGTGATTGAYG GGGTTGTWTGGTTTYGTTTA TTTGTGGTAT

(continued)

Table 1
(continued)

IGF2R (rs677882)	PCR primers	Amplified sequence position	Mg (mM)	Tp (°C)	Cycles	Pyrosequencing Primer	PCR Vol (μL)	Amplified sequence
Genotyping PCR	5'-Biotin-CCTGCCCTGG CCCTTGCTCT	chr6: 160426408–	0.5	61	50	5'-GTGGCCAA GGGGGAC	10	CCTGCCCTGGCCCTTGTCTGGGA GGCYGCRGTTCTCTGCGGCCT CTCGCYGGK
	5'-AAAAAGGCCCTGG AAGCG	160426599						GTCGCCCTTGGCCACTGGGCTG CGAGGAAGCCGGCGAGGAGT GGGTGTGCGCCCGCCCGGC TCTACCGCGAGGATGCAGT GGCCGCCAGCGTGGCCGTGC GGCTGGGTCACTCCCTTCCC GCTTCCAGGGCCCTTTT
Reference PCR	5'-TTTTGAAAAGGAGG TAGAAAAAGGT	chr6: 160426360–	1.6	58	50	5'-GGAAGGGAG TGATTTAGT 5'-ATTTTiyGT YGGTAGAG	10	TTTTGAAAAGGAGGTAGAAAA GGTTTGGAAAGCGGGAAGGG AGTGATTTAGTcGT ACGGTTACGTTGGCGGTATTG TATTTTCGTCGGTAGAGTC GGCGGGGCGTATATTTA TTTTCGTCGGTTTTTTCGTAG TTAGTGGTTAAGGGYGATW TYRGCAGAGGTCGTAGA GAATYGCRGTTTTTTAGATAAG GGTTAGGTAGGATTGYGGG TTGTCGTTATTACGTTAG GTTTTAGTTAGTTTTAAGA
	5'-Biotin-TCTTAAAAAC TAACTAAAAACCTAAC	160426616						
Note: reverse complementary strand after bisulfite is selected, where SNP= R								
Allele-specific PCR	5'-GYGGAAGGAGT GATTAG	chr6: 160426437–	1.2	63	33	5'-GGAAGGGA GTGATTAGT	20	GYGGAAGGAGTGATTAGTYGT AYGGTAYGTTGGYGTTATTGTAT TTTTYGTYGGTAGAGTYGGYGG GGYGTATATTTA
	5'-Biotin-CRATTCTCTA CRACCTCTCRC(C) ou 5'-Biotin-CRATTCTCTA CRACCTCTCRC(T)	160426584						TTTTTYGTGGTTTTTYGTAGTTT AGTGGTTAAGGGYGATWYR GCGAGAGGTCGTAGA GAATYG

2. In parallel, prepare the Pyrosequencing plate by diluting 4 pmol of the Pyrosequencing primer (*see* Table 1 for details) into 12 μ L of annealing buffer into the respective wells of the PSQ plate. Different Pyrosequencing primers can be used in different wells of the same plate.
3. Fill the four troughs of the vacuum preparation tool with 100 mL of 70 % ethanol, 100 mL of 0.2 M NaOH denaturing solution, 120 mL of washing buffer, and 150 mL of water (*see* Note 5).
4. Turn on the workstation to create a vacuum in the aspiration device (450 mmHg). Immerse the tips in water for several seconds for cleaning. Then, the PCR plate can be removed from the mixer and the binding mix can be aspirated. The beads remain on the filters of the tips.
5. After all binding mix has been aspirated; immerse the tips in ethanol 70 %, NaOH denaturing solution, and washing solution for 5, 5, and 10 s, respectively. Let the vacuum dry the beads for 10 s and turn it off above the PSQ plate.
6. Next, the filter tips can be immersed in the annealing mix of the PSQ plate in order to release the beads into the wells.
7. Heat the sequencing plate for 2 min at 80 °C on a thermoplate placed on a heating device for single DNA strand denaturation. Cool the plate at room temperature for 5 min for sequencing primer annealing.

3.2.3 Pyrosequencing Reaction

1. Create a new Pyrosequencing run on the Pyrosequencer using the PyroMark MD software. Indicate the Pyrosequencing assay used in the selected wells and the name and/or conditions of each well.
2. Fill the tips of the cartridge with the corresponding reagents and enzyme mix and dNTPs avoiding bubble formation. 180 μ L of reagents and enzyme mix is sufficient for a full plate and at least 150 μ L of each dNTPs should be used. The amount of dNTPs provided in the kit is not a limiting factor and therefore dNTPs can be used in excess.
3. Deposit a sealed Pyrosequencing test plate and the reagents' cartridge in the Pyrosequencer and perform the dispensation test to verify if the dispensing tips are working properly. Six homogeneous droplets should be clearly visible. If necessary tips should be changed.
4. Remove the test plate and put the cooled prepared Pyrosequencing plate and start the run. The length of a run is proportional to the number of dispensations (1 per min) and a genotyping run should not exceed 30 min.
5. After the end of the sequencing run, perform again a dispensation test to verify if some tips have become blocked during the run (*see* Note 6).

6. Analyze the results with the PyroMark MD software in Allele Quantification (AQ) mode, which permits the rapid identification of individuals heterozygous for the analyzed SNP(s) (*see Note 7*). To validate the Pyrogram[®] output of the Pyrosequencing reaction, it is necessary to have at least a signal intensity of 50 a.u. from the beginning until the end of the analyzed sequence for a peak corresponding to the incorporation of a single, nonvariable nucleotide.
7. Remove the cartridge and clean the tips with distilled water. Remaining reagents can be preserved at 4 °C and reused while dNTPs can be discarded.

3.3 Bisulfite Treatment

The protocol describes the translation of epigenetic into sequence differences using bisulfite conversion, which converts unmethylated cytosines into uracils (which will be replaced by thymines in the subsequent PCR reaction), while methylated cytosines are remain unchanged under the applied reaction conditions [24]. The protocol described here uses the EpiTect 96 Bisulfite Conversion kit, but a number of alternative kits are available and yield probably similar results. Although the analysis of allele-specific methylation is only informative for individuals heterozygous for a targeted SNP, a number of individuals homozygous for each genotype should be included as controls and to facilitate the set-up of the allele-specific amplification on bisulfite-treated DNA.

1. Dilute 500–1,000 ng of each DNA into a maximum of 20 µL pure water (*see Note 8*).
2. Add 120 mL of ethanol to the buffer BW (included in the EpiTect 96 kit).
3. Add 27 mL of ethanol to the buffer BD (included in the EpiTect 96 kit).
4. Add 1,350 µL of water to the lyophilized carrier RNA (included in the EpiTect 96 kit) and dissolve the RNA by vortexing.
5. Add 600 µL of the dissolved carrier RNA to each bottle of the BL buffer (included in the EpiTect 96 kit).
6. Immediately prior to use, add 9 mL H₂O to the Bisulfite Mix and vortex until the mix is dissolved. Small crystals may remain, do not wait too long before proceeding as the efficiency of the mix might be reduced.
7. Add to each sample in the Bisulfite Conversion Plate 85 µL of the diluted Bisulfite mix and 35 µL DNA Protect Buffer (included in the EpiTect 96 kit) and seal the plate and centrifuge briefly at 650×*g*. The color of the buffer will change from green to blue, indicating a sufficiently homogeneous mixture and a correct pH for the conversion reaction.

8. Incubate the plate in a thermocycler for 5 min at 95 °C, followed by 25 min at 60 °C, 5 min at 95 °C, 1 h 25 min at 60 °C, 5 min at 95 °C, 2 h 55 min at 60 °C, hold at 20 °C.
9. Centrifuge the conversion plate briefly at $650\times g$.
10. Place an EpiTect plate onto the provided S-Block and dispense 560 μL BL buffer containing the carrier RNA into each well.
11. Transfer the complete bisulfite reactions to the EpiTect Plate with the BL buffer and mix gently by pipetting four times up and down.
12. Seal the plate and centrifuge the plate with the S-Block at $5,800\times g$ for 1 min. Discard the flow-through.
13. Remove the sealing and add carefully 250 μL BD buffer to each sample, seal again, and incubate at room temperature for 15 min.
14. Centrifuge at $5,800\times g$ for 1 min.
15. Remove the sealing and add 500 μL BD buffer to each sample, seal again, centrifuge at $5,800\times g$ for 1 min, and discard the flow-through.
16. Repeat **step 15**.
17. Remove the sealing and add carefully 250 μL ethanol to each sample, seal again, centrifuge at $5,800\times g$ for 1 min, and discard the flow-through.
18. Dispose of the S-block, place the EpiTect 96 plate on top of an EpiTect Elution Plate, remove the sealing, and centrifuge at $5,800\times g$ for 15 min at 40 °C.
19. Place the EpiTect 96 plate on top of a new EpiTect Elution Plate and add 70 μL of the Elution buffer (EB) onto the center of the membrane.
20. Elute the DNA by centrifugation at $5,800\times g$ for 1 min.

3.4 Quantification of Bisulfite-Treated DNA (Optional)

The protocol assesses the quantity and integrity of DNA recovered from the bisulfite conversion amplifying an *ALU* repeat consensus sequence in a bisulfite dependent, but methylation-independent manner. The protocol described in this paragraph is based on a publication describing quality control experiments for DNA methylation analysis by MethyLight [25], the probe sequence has been modified to incorporate LNAs instead of the proprietary MGB—quencher to ensure specificity.

1. Prepare a serial dilution of (commercial) bisulfite-treated DNA with a known concentration ranging from 10 to 0.01 ng/ μL .
2. Add 1 $\times$ of probe MasterMix, 0.1 μM probe, 0.3 μM of forward and reverse primer, and 3 μL of bisulfite-converted DNA (sample or standard) to a final volume of 20 μL .

3. PCR cycling conditions include an initial denaturation step for 3 min at 95 °C, followed by 50 cycles of 3 s denaturation at 95 °C and 30 s annealing, probe hydrolysis, and elongation at 60 °C.
4. Acquire fluorescence at the end of the hydrolysis step in each cycle.
5. Determine the amount of amplifiable bisulfite-converted DNA from the C_p -values generated by the instrument's software.

3.5 Analysis of the DNA Methylation Patterns of a Target Region

In this section the DNA methylation analysis of a target region independent of the presence of heterozygous SNPs is described. This step allows verifying that DNA methylation patterns will be informative and potentially exclusively (methylation degree of ~50 %) or preferentially (methylation degree of 20–50 %) associated with a single allele. If the analyzed region surrounding the selected SNP (*see* Subheading 3.2) is completely unmethylated or fully methylated, alternative genetic variation should be investigated. For a detailed description of DNA methylation analysis by Pyrosequencing, *see* also Chapter 13 in this volume.

1. Determine the optimal annealing temperature for the amplification of the target sequence designed in Subheading 3.1.2. This is performed on a gradient thermocycler using bisulfite-treated DNA as template.
2. Amplify the region of interest using 10–15 ng of bisulfite-converted DNA, 1× HotStar Taq buffer supplemented with 1.6 mM MgCl₂, 0.1 mM of each dNTP, 2.0 U of HotStar Taq polymerase, and 200 nM of forward and reverse primers in a total volume of 25 µL.
3. Typical PCR cycling conditions include an initial denaturation step performed 15 min at 95 °C, followed by 50 cycles of 30 s denaturation at 95 °C, 30 s of a gradient of annealing temperatures ranging from 50 to 70 °C, and 15 s elongation at 72 °C. Final extension is performed for 5 min at 72 °C.
4. Deposit 5 µL of PCR product supplemented with 1× loading dye as well as a size standard on a 2 % agarose gel including a fluorescent dye (example: ethidium bromide). Perform a horizontal electrophoresis in 1× TBE at 5 V/cm for 1 h. The specificity of the amplifications is assessed by the evaluation of the length of the PCR product and the highest temperature allowing a strong and specific amplification product is selected as the optimal annealing temperature.
5. Amplify the region of interest using 10–15 ng of bisulfite-converted DNA, 1× HotStar Taq buffer supplemented with 1.6 mM MgCl₂, 0.1 mM of each dNTP, 2.0 U of HotStar Taq polymerase, and 200 nM of forward and reverse primers in a total volume of 25 µL.

6. PCR cycling conditions include an initial denaturation step for 15 min at 95 °C, followed by 50 cycles of 30 s denaturation at 95 °C, 30 s at the optimal annealing temperature defined in **steps 2–4** of this section (*see* Table 1 for examples), and 15 s elongation at 72 °C. Final extension is performed for 5 min at 72 °C.
7. Verify 5 µL of the PCR products on a 2 % agarose electrophoresis gel as described in **step 4**.
8. PCR products can be stored at +4 °C for several days or at –20 °C for several months.
9. Perform sample preparation and Pyrosequencing analysis as described in Subheadings 3.2.2 and 3.2.3 with the exception that the Q-CpG software is used to enter the Pyrosequencing assays, start and control the Pyrosequencing run, and analyze quantitatively the results.

3.6 Analysis of the Allele-Specific DNA Methylation Patterns of a Target Region

Following the identification of the presence of heterozygous genetic variation (Subheading 3.2) and the exclusion of complete absence of DNA methylation or complete DNA methylation using standard Pyrosequencing (Subheading 3.5), this section describes the set-up and optimization of allele-specific amplification on bisulfite-treated DNA using a heterozygous SNP and two allele-specific primers for the selective amplification of a single allele. As the bisulfite-treated DNA displays a lower genomic complexity, the set-up of these PCRs do require some optimization and the availability of reference samples representing the three genotypes is required for the successful optimization and validation of the allele-specific assay. The use of LNAs at the last base of one of the amplification primers facilitates greatly the allele discrimination.

1. Determine the optimal annealing temperature for the amplification of each allele target sequence designed in Subheading 3.1.3. This is performed on a gradient thermocycler using DNA heterozygous for the targeted SNP as template.
2. Amplify the region of interest using 10–15 ng of bisulfite-converted DNA, 1× HotStar Taq buffer supplemented with 1.0–1.2 mM MgCl₂, 0.1 mM of each dNTP, 2.0 U of HotStar Taq polymerase, and 200 nM of forward and reverse primers in a total volume of 25 µL.
3. Typical PCR cycling conditions include an initial denaturation step performed for 15 min at 95 °C, followed by 50 cycles of 30 s denaturation at 95 °C, 30 s annealing at a gradient of annealing temperatures ranging from 50 to 70 °C, and 15 s elongation at 72 °C. Final extension is performed for 5 min at 72 °C.
4. Deposit 5 µL of PCR product supplemented with 1× loading dye as well as a DNA size standard on a 2 % agarose gel including

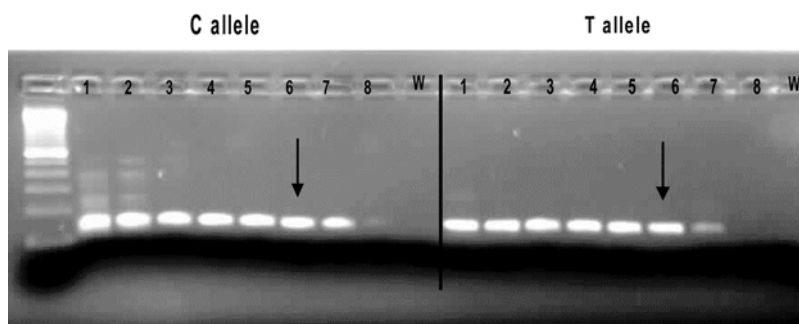


Fig. 1 Gradient for the selection of the optimal annealing temperature for the amplification of a target region in *IGF2R*. The selected temperature corresponding to the highest temperature where there is a strong amplification for both alleles is highlighted with an *arrow*. Temperature gradient: Well 1: 51.4 °C, Well 2: 53.2 °C, Well 3: 54.5 °C, Well 4: 58.1 °C, Well 5: 60.8 °C, Well 6: 63.5 °C, Well 7: 64.8 °C, Well 8: 66.6 °C

- a fluorescent dye (example: ethidium bromide). Perform a horizontal electrophoresis in 1× TBE at 5 V/cm for 1 h.
5. Select the highest temperature allowing a strong and specific amplification product as the optimal annealing temperature (Fig. 1).
6. Verify the allele-specificity of the amplification by analyzing the three different genotypes at the selected temperature.
7. Amplify the region of interest using 10–15 ng of bisulfite-converted DNA, 1× HotStar Taq buffer supplemented with 1.0–1.2 mM MgCl₂, 0.1 mM of each dNTP, 2.0 U of HotStar Taq polymerase, and 200 nM of forward and reverse primers in a total volume of 25 µL.
8. The PCR programs consists of a denaturation step at 95 °C for 15 min followed by 33–35 cycles of 30 s at 95 °C, 30 s at the correct annealing temperature (*see* Table 1 for details), 15 s at 72 °C, and 5 min of final extension at 72 °C.
9. Deposit 5 µL of PCR product supplemented with 1× loading dye as well as a DNA size standard on a 2 % agarose gel including a fluorescent dye (example: ethidium bromide). Perform a horizontal electrophoresis in 1× TBE at 5 V/cm for 1 h. The results show a preferential amplification of the complementary allele (Fig. 2). However, weak signals of the mismatched allele can be visible and might result in Pyrograms of slightly reduced but still evaluable quality.
10. Perform sample preparation and Pyrosequencing analysis as described in Subheadings 3.2.2 and 3.2.3 with the exception that the Q-CpG software is used to enter the Pyrosequencing assays, execute the Pyrosequencing run, and analyze quantitatively the results.

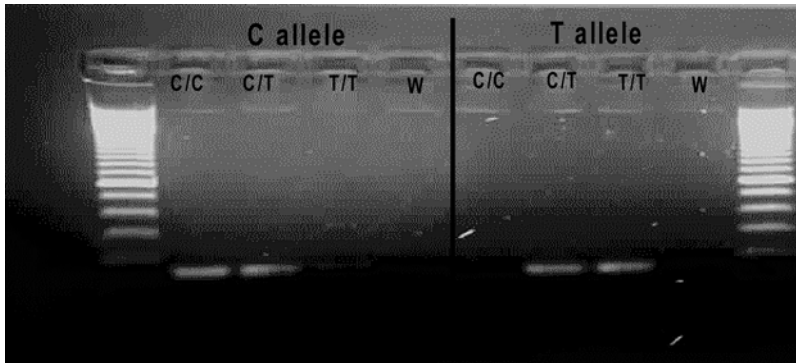


Fig. 2 Validation of the selected temperature on all three possible genotypes for *IGF2R*. The targeted allele is preferentially amplified, but weak by-products of the other allele are still visible at 35 cycles and lead to readable Pyrograms. Reduction of the amplification to 33 cycles yields completely specific amplification products

11. Verify the allele-specificity in the Pyrograms. If “high quality” Pyrograms are obtained for the two homozygous genotypes with the same allele-specific primer, indicating that the amplification was not sufficiently allele-specific, the allele selectivity needs to be further improved.
12. Repeat **steps 6–11** reducing the amount of magnesium added and/or the number of PCR cycles performed and verify the allele-specificity by Pyrosequencing of the resulting amplification products until alleles are selectively amplified (Fig. 3) (*see Notes 9 and 10*).
13. For further quality control of the allele-specific amplification system, verify that the average of the methylation degree of the two allele-specific DNA methylation products corresponds to the mean methylation degree of the same CpGs analyzed using a non-allele-specific amplification as described in Subheading 3.5 (*see Table 2*).
14. The volume of the PCR amplification allows for the analysis of multiple CpGs in the amplification product using two or three Pyrosequencing primers. If the volume is not sufficient or the Pyrosequencing reaction fails due to a technical problem, refer to **Note 6**.

4 Notes

1. As the two strands of genomic DNA are no longer complementary after bisulfite treatment, both strands can be analyzed for possible amplification products. MethPrimer takes only the forward strand into account and the reverse strand has to be created manually or using one of the variety of software tools.

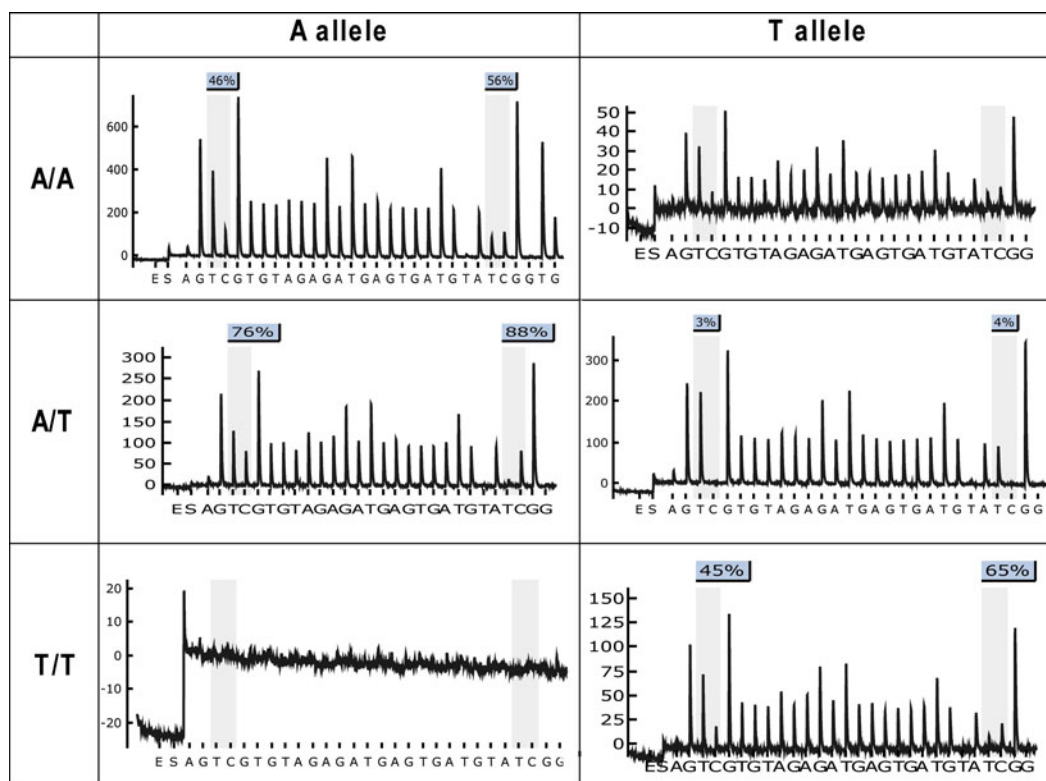


Fig. 3 Pyrograms for three samples representing the three possible genotypes for *PEG3*. The signal for the targeted allele is strong and specific for the targeted alleles and the allele noncomplementary to the LNA-modified amplification primer yields only very weak signals corresponding to the Pyrosequencing background. Heterozygous individuals show the expected allele-specific DNA methylation patterns with high methylation levels on one allele and absence of DNA methylation on the other allele

- Pyrosequencing primers can also be designed using the Pyrosequencing Assay Design software. The software has not been designed for DNA methylation analysis and the assay set-up on bisulfite-treated DNA leads to low scores for many assays. Despite the low score, many assays do work in practice. For simplification, enter the bisulfite-treated sequence of the amplification product into the Pyrosequencing software. Change the first targeted “CpG” to “YG” so that the software creates a primer just before this “artificial SNP.” Repeat this for several of the CpGs to ensure a complete coverage of the amplification product.
- In our experience the modification of more than the 3' base with a LNA has no positive effect on the capacity of the primer for allele differentiation. Similarly, experiments using primers with a mismatch at the third base from the 3' terminus to destabilize the primers for both alleles [26] did in general not improve the allele discrimination.

Table 2

Comparison of the mean DNA methylation levels obtained by allele-specific PCR and standard PCR on bisulfite-treated DNA analyzing *IGF2*

Difference of methylation (%)	Number of samples					Overall analyzed CpGs
	CpG 1	CpG 2	CpG 3	CpG 4	CpG 5	
≤ -10 %	8	6	1	2	8	3
Between -10 and -5 %	30	17	17	9	22	15
Between -5 and 5 %	51	69	73	72	59	73
Between 5 and 10 %	2	1	1	8	4	1
>10 %	3	1	2	3	1	2
Total samples	94					

The distribution of differences between DNA methylation data obtained by the analysis of the overall DNA methylation patterns (without allele-specific amplification) and the average percentage of both allele-specific PCRs on 94 samples from the CEPH reference collection. Data is given for each of the analyzed CpGs as well as the overall methylation level calculated as mean of all CpGs for an individual. Most samples have a difference of less than 5 % between the two protocols

4. It is recommended to analyze for each Pyrosequencing assay in parallel the PCR product without a Pyrosequencing primer, the negative (no template) control of the PCR amplification and the Pyrosequencing primer without any PCR product to ensure the absence of contaminating amplification product or signals due to primer interactions.
5. A greater volume of washing buffer should be put in the troughs compared to the troughs containing ethanol and NaOH to assure the complete removal of any traces of NaOH, which might inhibit the downstream Pyrosequencing reactions.
6. If not sufficient PCR product is available to perform all Pyrosequencing reactions on a target region amplified in a single PCR reaction or if a problem with the nucleotide dispensation was encountered, it is possible to recycle the template strand for additional analyses as the biotinylated strand is not altered during Pyrosequencing. In case of dispensation problems during the run, this method also allows to repeat the Pyrosequencing run without the need to redo the PCR amplification and use up more of the sample DNA. Although there is a slight loss in intensity due to incomplete recovery of the biotinylated strand, quantitative results are unaltered for several cycles of Pyrosequencing on the same template [27].
 - (a) Prepare a Pyrosequencing plate with a new Pyrosequencing primer and the workstation as described in Subheading 3.2.2.
 - (b) Add 20 µL binding buffer to the completed sequencing reaction (from Subheading 3.5 or 3.6) and resuspend the Sepharose beads by vigorous pipetting.

Purify the mixture without further incubation as described in **steps 4–7** of Subheading **3.2.2**.

7. The Allele Quantification mode allows assessing the presence of potential preferential amplification of one allele. If allele-frequencies deviating significantly from 0, 50, or 100 % for one allele are detected, the assay should be redesigned as amplification of nontargeted regions with high homology occurs or unknown genetic variations underlying the amplification primers are present.
8. The amount of input DNA can be decreased to 100 ng or even less using adapted protocols by the manufacturer. However, it should be kept in mind that for the accurate and reproducible quantification of the DNA methylation degree by Pyrosequencing a minimal amount of 10 ng for the input DNA is required [9]. The protocol presented in the main text describes the conversion of high molecular weight DNA. Specialized protocols for the conversion of fragmented DNA such as DNA extracted from FFPE blocks are provided by the manufacturer.
9. Reducing the number of amplification cycles has in most cases a greater impact on the allele-specificity than variation of the magnesium concentration. For alleles with a temperature increasing variant (G or C) a slight increase in the annealing temperature might also be beneficial.
10. In many cases the protocol for the allele-specific amplification will differ between the two alleles in several parameters including the number of cycles and/or the optimal annealing temperature (see, e.g., *IGF2* in Table 1) thus requiring separate amplification reactions for both alleles. If a limited number of samples is analyzed, it is preferable to process both amplification products of the same samples together during the sample preparation and the Pyrosequencing reaction to reduce inter-plate/run variation in the accurate estimation of the DNA methylation degree (see also Chapter 4).

References

1. Tost J (2010) DNA methylation: an introduction to the biology and the disease-associated changes of a promising biomarker. *Mol Biotechnol* 44:71–81
2. Kerkel K, Spadola A, Yuan E et al (2008) Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat Genet* 40:904–908
3. Schalkwyk LC, Meaburn EL, Smith R et al (2010) Allelic skewing of DNA methylation is widespread across the genome. *Am J Hum Genet* 86:196–212
4. Hutchinson JN, Raj T, Fagerness J et al (2014) Allele-specific methylation occurs at genetic variants associated with complex disease. *PLoS One* 9:e98464
5. Ronaghi M, Uhlen M, Nyren P (1998) A sequencing method based on real-time pyrophosphate. *Science* 281(363):365
6. Harrington CT, Lin EI, Olson MT et al (2013) Fundamentals of pyrosequencing. *Arch Pathol Lab Med* 137:1296–1303
7. Langae T, Ronaghi M (2005) Genetic variation analyses by pyrosequencing. *Mutat Res* 573:96–102

8. Tost J, Dunker J, Gut IG (2003) Analysis and quantification of multiple methylation variable positions in CpG islands by pyrosequencing. *Biotechniques* 35:152–156
9. Dupont JM, Tost J, Jammes H et al (2004) De novo quantitative bisulfite sequencing using the pyrosequencing technology. *Anal Biochem* 333:119–127
10. Tost J, Gut IG (2007) DNA methylation analysis by pyrosequencing. *Nat Protoc* 2:2265–2275
11. Ronaghi M (2001) Pyrosequencing sheds light on DNA sequencing. *Genome Res* 11:3–11
12. Ahmadian A, Ehn M, Hober S (2006) Pyrosequencing: history, biochemistry and future. *Clin Chim Acta* 363:83–94
13. Ogino S, Kawasaki T, Brahmandam M et al (2005) Sensitive sequencing method for KRAS mutation detection by pyrosequencing. *J Mol Diagn* 7:413–421
14. Lundin KE, Hojland T, Hansen BR et al (2013) Biological activity and biotechnological aspects of locked nucleic acids. *Adv Genet* 82:47–107
15. How Kit A, Mazaleyra N, Daunay A et al (2013) Sensitive detection of KRAS mutations using enhanced-ice-COLD-PCR mutation enrichment and direct sequence identification. *Hum Mutat* 34:1568–1580
16. Peters J (2014) The role of genomic imprinting in biology and disease: an expanding view. *Nat Rev Genet* 15:517–530
17. Wong HL, Byun HM, Kwan JM et al (2006) Rapid and quantitative method of allele-specific DNA methylation analysis. *Biotechniques* 41: 734–739
18. Shaw RJ, Akufo-Tetteh EK, Risk JM et al (2006) Methylation enrichment pyrosequencing: combining the specificity of MSP with validation by pyrosequencing. *Nucleic Acids Res* 34:e78
19. Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23:1289–1291
20. Untergasser A, Cutcutache I, Koressaar T et al (2012) Primer3: new capabilities and interfaces. *Nucleic Acids Res* 40:e115
21. Li LC, Dahiya R (2002) MethPrimer: designing primers for methylation PCRs. *Bioinformatics* 18:1427–1431
22. Zhou GH, Gotou M, Kajiyama T et al (2005) Multiplex SNP typing by bioluminometric assay coupled with terminator incorporation (BATI). *Nucleic Acids Res* 33:e133
23. Aranyi T, Varadi A, Simon I et al (2006) The BiSearch web server. *BMC Bioinformatics* 7:431
24. Frommer M, McDonald LE, Millar DS et al (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci U S A* 89:1827–1831
25. Campan M, Weisenberger DJ, Trinh B et al (2009) MethyLight. *Methods Mol Biol* 507: 325–337
26. Okimoto R, Dodgson JB (1996) Improved PCR amplification of multiple specific alleles (PAMSA) using internally mismatched primers. *Biotechniques* 21:20–22
27. Tost J, El abdalaoui H, Gut IG (2006) Serial pyrosequencing for quantitative DNA methylation analysis. *Biotechniques* 40:721–726

Free ebooks ==> www.Ebook777.com

Chapter 22

DNA Methylation Analysis of ChIP Products at Single Nucleotide Resolution by Pyrosequencing®

Céline Moison, Fanny Assemat, Antoine Daunay,
Paola B. Arimondo, and Jörg Tost

Abstract

Interaction and co-occurrence of protein and DNA-based epigenetic modifications have become a topic of interest for many fundamental and biomedical questions. We describe within this chapter a protocol that combines two techniques in order to determine the methylation status of the DNA specifically associated with a protein of interest. First, DNA that directly interacts with the selected protein (such as a specific histone modification, a transcription factor, or any other DNA-associated protein) is purified by standard chromatin immunoprecipitation (ChIP). Second, the level of DNA methylation of this immunoprecipitated DNA is measured by bisulfite conversion and Pyrosequencing, a quantitative sequencing-by-synthesis method. This procedure allows determining the methylation status of genomic DNA associated to a specific protein at single nucleotide resolution.

Key words ChIP, DNA methylation, Pyrosequencing®, Bisulfite conversion, Immunoprecipitation

1 Introduction

Epigenetic modifications such as posttranscriptional histone modifications, histone variants, and DNA methylation lay the regulatory landscape for the gene expression program of a cell. These diverse molecular mechanisms were all found to be closely intertwined and stabilize each other to ensure the faithful propagation of an epigenetic state over time and especially through cell division. While many technologies have been developed for the comprehensive analysis of a single type of epigenetic modification, few can address the co-occurrence and interaction of different modifications [1–3].

Chromatin immunoprecipitation (ChIP) is a well-established method in cellular biology to study the specific interaction between a protein of interest and genomic DNA [4]. For example, this technique has been extensively used to identify transcription factor binding sites. In cell lines or tissue samples, ChIP allows a snap

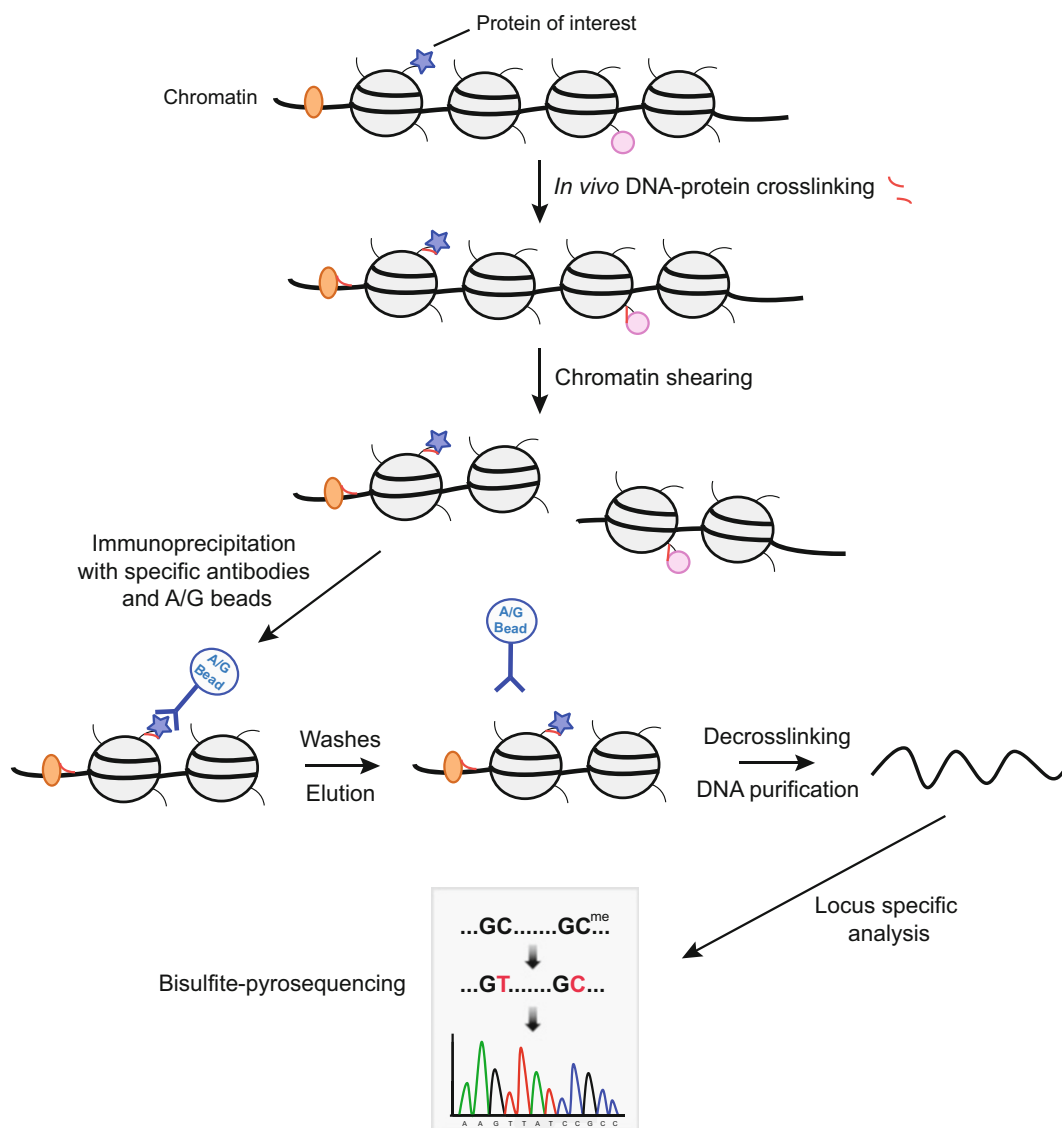


Fig. 1 ChIP-bisulfite Pyrosequencing protocol. Schematic representation of the multistep protocol of the ChIP and the Pyrosequencing analysis after bisulfite conversion. Nucleosomes are represented with *circles* wrapped by DNA (*black line*), histone modifications and DNA-associated proteins are depicted in *coloured forms*. The antibody is represented as a *blue Y*

shot of the chromatin at a specific time point and to study DNA-associated proteins at a locus of interest or genome-wide.

The outline of a ChIP experiment is presented in Fig. 1.

Briefly, proteins are crosslinked to the DNA by chemicals in order to conserve the in vivo chromatin architecture. Chromatin is extracted and randomly fragmented by sonication into 200–600 base pair fragments. Then, DNA–protein complexes are immunoprecipitated using a specific antibody and protein A/G agarose resin. Finally, covalent crosslinks are reversed by heating and DNA

is purified after RNase A and Proteinase K treatment. At this point, a small amount of purified DNA is available and several options exist to analyze the DNA associated to the protein of interest: PCR or qPCR allows the detection and quantification of the analyzed modification at a locus of interest by using specific primers, while microarrays (ChIP-on-chip) or next generation sequencing (ChIP-seq) provide a genome-wide picture (i.e., identification of all binding sites of a transcription factor; mapping of an histone modification on the entire genome) [3, 4].

Since the DNA can also be chemically modified and, in particular, the human genome can be methylated at position 5 of cytidines at CpG dinucleotides, we developed a protocol to analyze the DNA methylation levels of immunoprecipitated DNA at a single nucleotide resolution using bisulfite conversion and analysis of the ChIPed DNA by Pyrosequencing®. Pyrosequencing is a quantitative real-time sequencing method that is also frequently used for the analysis of DNA methylation patterns [5–7] (*see also* Chapters 13–16). Pyrosequencing is based on the presence or absence of the incorporation of a nucleotide during primer extension. In contrast to Sanger sequencing, which relies on the random incorporation of fluorescent ddNTPs during primer extension steps, only one specific nucleotide is present at any time in the Pyrosequencing reaction. Following nucleotide incorporation, the released pyrophosphate (PPi) is used as substrate in combination with adenosine 5' phosphosulfate (APS) by an ATP sulfurylase to produce ATP [8] (*see also* Chapter 1 for a detailed description of the biochemistry). The latter is in turn used by luciferase to oxidize luciferin into oxyluciferin resulting in a light emission proportional to the amount of incorporated nucleotide. DNA methylation analysis by Pyrosequencing thus permits simultaneous analysis and quantification of the methylation status of several CpG positions in close proximity. This point is of particular interest as successive CpGs might display significantly different levels of methylation as demonstrated in the differentially methylated region of imprinting genes (*see also* Chapters 18 and 19) as well as at promoters devoid of a CpG island. Pyrosequencing combines the advantages of sequence-based approaches, such as in-built quality control and resolution of individual CpG sites, with the possibilities of medium to high throughput (for DNA methylation analysis) and the advantages of PCR-based technologies. It features a limit of detection of ~5 % for the minor component of a quantitative signal and a quantitative resolution of 5 % or better (*see* Chapter 4 for a detailed discussion of the limitations of quantification and detection).

The combination of ChIP with the Pyrosequencing technique enables to determine at a given genomic locus if a protein is bound to methylated DNA or unmethylated DNA. This approach permits thus studying a putative methyl binding protein or the association of specific histone modification with methylated or nonmethylated DNA. We previously used this approach to demonstrate that in

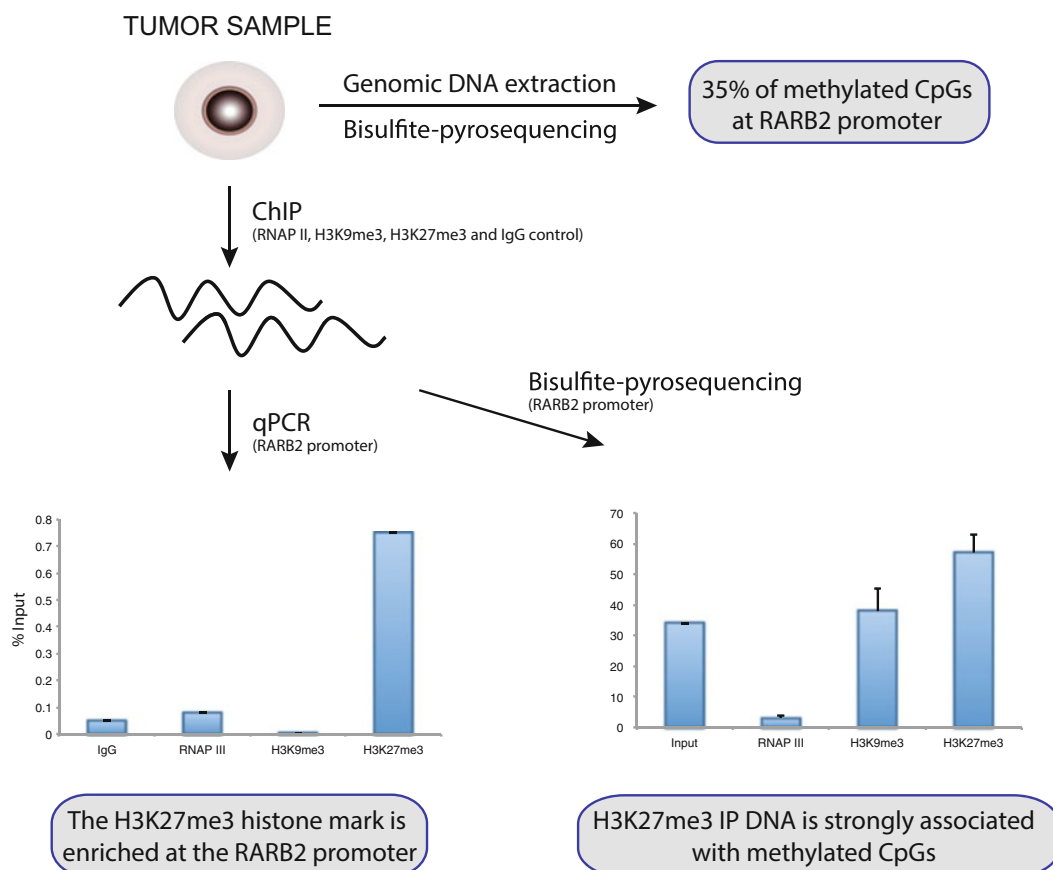


Fig. 2 Coexistence of DNA methylation and histone marks at a tumor suppressor gene in prostate tumors. Genomic DNA extracted from a tumor sample of a prostate cancer patient was bisulfite-converted and analyzed by Pyrosequencing (10 CpGs) showing an average DNA methylation of 35 % at the *RARB2* promoter. ChIP-qPCR analysis showed that *RARB2* promoter is enriched in H3K27me3 polycomb histone mark, while no enrichment in RNAP II is detected. Mouse IgG antibody is used as a negative control of chromatin IP and displays background enrichment. Enrichments are expressed as percentage of input. Purified DNA from ChIP experiment was also treated by bisulfite and analysed by Pyrosequencing to measure the DNA methylation level and revealed the coexistence of the studied histone modifications and DNA methylation. As expected, input ChIP sample showed a similar DNA methylation level as genomic DNA. No DNA methylation was detected in the RNAP II enriched DNA sample compared to H3K9me3 and H3K27me3 [9]. We can conclude that these repressive histone marks coexist with DNA hypermethylation at the *RARB2* promoter and is consistent with the lack of expression of the tumour suppressor gene in prostate tumours.

prostate tumor samples the DNA methylation of the *RARB2* promoter is associated with two repressive histone modifications (Fig. 2, [9]).

We describe herein a ChIP protocol that we applied recently on cell lines and human tumor tissue samples (Fig. 1). However numerous other protocols exist in the literature [10–16] to generate immunoprecipitated fragments of DNA and can also be used to analyze DNA methylation levels in combination with the bisulfite Pyrosequencing protocol described hereafter. Of note, while the

protocol described in the chapter focuses on the analysis of specific loci of interest such as a promoter or another gene regulatory elements, other studies have combined ChIP with genome-wide sequencing methods to obtain a more comprehensive overview of the DNA methylation status with specific histone modifications [17, 18].

2 Materials

2.1 *In Vivo* Crosslinking and Chromatin Extraction

1. Formaldehyde 37 %.
2. Glycine 1 M.
3. 1× Dulbecco's Phosphate Buffered Saline (PBS).
4. Protease inhibitor cocktail (e.g., Roche, Complete Tablets).
5. Cell scraper.
6. Buffer 1: 50 mM Hepes-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 10 % glycerol, 0.5 % NP-40, 0.25 % Triton X-100.
7. Buffer 2: 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 10 mM Tris-HCl pH 8.0.
8. Buffer 3: 50 mM Tris-HCl pH 8.0, 0.1 % SDS, 0.95 % NP-40, 0.1 % Na-DOC, 10 mM EDTA, 150 mM NaCl.
9. Tubes (1.5 and 2 mL).
10. Centrifuge.
11. Rotating wheel.

All buffers should be stored at +4 °C.

2.2 *Chromatin* Fragmentation

1. Sonicator and equipment.
2. Tubes (polystyrene conical, 15 mL).
3. Agarose gel and gel electrophoresis apparatus.

2.3 *Immunopre-* *cipitation*

1. Protein A/G UltraLink Resin (20 mg/mL).
2. Bovin Serum Albumin (BSA) 1 mg/mL.
3. Quant-iT ds DNA broad-range assay kit.
4. Antibodies: control and specific to the protein(s) of interest.
5. Low-Salt Buffer: 20 mM Tris-HCl pH 8, 0.1 % SDS, 1 % Triton X-100, 2 mM EDTA, 150 mM NaCl.
6. High-Salt Buffer: 20 mM Tris-HCl pH 8, 0.1 % SDS, 1 % Triton X-100, 2 mM EDTA, 500 mM NaCl.
7. Li Buffer: 0.25 mM LiCl, 1 % NP-40, 1 % Na-DOC, 10 mM Tris-HCl pH 8, 1 mM EDTA.
8. TE Buffer: 50 mM Tris-HCl pH 8, 50 mM NaCl, 1 mM EDTA.

All buffers and reagents should be stored at +4 °C.

2.4 DNA Purification

1. Elution Buffer: 50 mM Tris-HCl pH 8, 1 mM EDTA, 1 % SDS.
2. Thermomixer.
3. NaCl 5 M.
4. RNase A (0.5 mg/mL).
5. Proteinase K (20 mg/mL).
6. Phenol:chloroform:isoamylalcohol (25:24:1).
7. Phase Lock Gel Light 1.5 mL.
8. Ethanol (70 and 100 %).
9. Sodium Acetate, 3 M, pH 5.
10. Glycogen (5 mg/mL).
11. Nuclease free water.
12. Primers for PCR/qPCR and apparatus.

All buffers should be stored at +4 °C; RNase A and proteinase K at -20 °C.

2.5 Bisulfite Conversion

1. EZ DNA Methylation-Gold™ (Zymo Research).
2. Ethanol (100 % PCR grade).
3. Mixed Human DNA (e.g., Promega).

2.6 Quantification of Amplifiable DNA

1. Commercial fully unmethylated or methylated DNA (e.g., Qiagen).
2. Probe MasterMix (Roche).
3. Probe for real-time PCR amplification 5'-6FAM- CCT ACC TTA+ AC+ C T+ C+ C C- BBQ, modified from [19]; (BBQ= Black Berry Quencher, += LNA modified nucleotide).
4. Forward (5'-GGT TAG GTA TAG TGG TTT ATA TTT GTA ATT TTA GTA) and reverse (5'-ATT AAC TAA ACT AAT CTT AAA CTC CTA ACC TCA) primer for real-time PCR amplification [19].
5. 96-well real-time Thermocycler with associated software (e.g., LightCycler480, Roche Molecular Diagnostics).
6. Plates for real-time thermocycler.

2.7 PCR Amplification

1. Primers for PCR amplification.
2. Biotinylated PCR primers.
3. HotStar Taq DNA Polymerase.
4. dNTPs.
5. Skirted 96-well plates.
6. Thermocycler.

All reagents should be stored at -20 °C.

2.8 Gel Electrophoresis

1. TBE 10×.
2. UltraPure™ Agarose.
3. 100 bp DNA ladder.
4. 6× DNA Loading Dye Buffer Blue.
5. Gel molds and combs.
6. Electrophoresis apparatus.
7. Gel electrophoresis imaging system.

2.9 Template Preparation for Pyrosequencing

1. 96-well Low Profile PCR Plate, Skirted.
2. Streptavidin Sepharose HP beads.
3. PyroMark® Q96 Vacuum Workstation (220 V).
4. PyroMark Q96 Plate Low.
5. PyroMark Q96 Sample Prep Thermoplate Low.
6. PyroMark Q96 Vacuum Prep Trough.
7. Binding buffer: 10 mM Tris-HCl, 2 M NaCl, 1 mM EDTA, 0.1 % Tween 20, pH 7.6.
8. Denaturing solution: 0.2 M NaOH.
9. Washing buffer: 10 mM Tris-acetate, pH 7.6.
10. Annealing buffer: 20 mM Tris-acetate, 2 mM Mg-acetate, pH 7.6.
11. Costar microplate sealing tape.
12. Thermomixer or similar (room temperature).
13. Pyrosequencing primer.

All reagents used for this step should be stored at room temperature except for the streptavidin-coated Sepharose beads (+4 °C) and the Pyrosequencing primer (−20 °C).

2.10 Pyrosequencing Analysis

1. PyroMark Dispensing Tip Holder.
2. PyroMark Reagent and Nucleotide Dispensing Tip.
3. PyroMark Q96 MD Pyrosequencer.
4. PyroMark Gold SQA Reagents (1×96).
5. Q-CpG software.

3 Methods**3.1 In Vivo Crosslinking and Chromatin Extraction**

1. For adherent cells in a 150 cm² dish at 80 % confluence (around 1×10⁷ cells), add 37 % formaldehyde directly to the media at a final concentration of 1 %, gently swirl dish to mix, and incubate 10 min at room temperature (RT) (*see* **Notes 1–3**).

2. Add freshly made glycine to a final concentration of 0.125 M to stop crosslinking and incubate 5 min at RT. Media will turn yellow due to pH change.
3. Place dish on ice. Wash cells twice with 20 mL of cold 1× PBS.
4. Add 2 mL of cold 1× PBS complemented with a protease inhibitor cocktail and collect cells by scraping. Transfer samples into 2 mL tubes.
5. Centrifuge at 4 °C for 5 min at 1,000×*g* to pellet cells.
6. Discard the supernatant and suspend cell pellet in 1 mL of cold Buffer 1 (1 mL per dish). Rock on a wheel at 4 °C for 10 min.
7. Centrifuge at 1,000×*g* for 5 min at 4 °C to collect nucleus.
8. Discard the supernatant and suspend cell pellet in 1 mL of cold Buffer 2 to lyse the nuclear membrane. Rock on a wheel at 4 °C for 10 min.
9. Centrifuge at 4 °C for 5 min at 1,000×*g*.
10. Discard the supernatant and suspend cell pellet in 1.3 mL of cold Buffer 3.

3.2 Chromatin Fragmentation

1. Sonicate-extracted chromatin to generate DNA fragments of 200–600 bp (*see Note 4*). We typically use a Bioruptor from Diagenode set to HIGH intensity with cycles of 30 s ON—30 s OFF during 5–30 min. Sonication time is critical and needs optimization for each cell line or tissue sample (*see Note 5*).
2. Transfer samples in prechilled 15 mL polystyrene conical tube and perform sonication at 4 °C (put ice in the water bath of the sonicator before use and, if needed, during sonication).
3. Centrifuge at 4 °C for 10 min at 8,000×*g* and transfer the supernatant containing the sheared chromatin to a 1.5 mL tube. Samples can be divided in aliquots and frozen at –80 °C at this stage.

3.3 Immunoprecipitation

3.3.1 Preclearing

1. Preclear total sonicated chromatin (around 1.2 mL) with 150 µL of unblocked protein A/G beads during 2 h at 4 °C on a rotating wheel.
2. Centrifuge at 4 °C for 10 min at 1,000×*g* and preserve the supernatant that contains precleared chromatin.
3. Quantify the chromatin with Quant-It dsDNA Assay kit High Sensitivity following the manufacturer's protocol. 500 ng of chromatin is needed for each IP.
4. At this stage, save 10 % (50 ng) of the precleared chromatin that will be used as your INPUT control. Keep at 4 °C until

use (Subheading 3.4 DNA purification). It represents the total chromatin on which IP will be performed and allows distinguishing background from specific signal.

3.3.2 Blocking of the Beads

1. Pipet 200 μ L of the Protein A/G beads and wash three times with 1 mL of Buffer 3 by cycles of centrifugation (3 min, 1,000 $\times g$, 4 °C) and incubation on a rotating wheel (5 min, 4 °C).
2. Preblock the protein A/G beads in 1 mL of Buffer 3 containing 1 mg/mL BSA on a rotating wheel overnight at 4 °C.
3. Wash three times the blocked protein A/G beads in 1 mL of Buffer 3 by cycles of centrifugation (3 min, 1,000 $\times g$, 4 °C) and incubation on a rotating wheel (5 min, 4 °C). After the final wash, resuspend the protein A/G beads in 1 mL of Buffer 3. Beads can be aliquoted and stored at 4 °C.

3.3.3 Incubation with Antibodies

1. 500 ng of precleared chromatin is needed for each immunoprecipitation. Dilute chromatin to a final volume of 400 μ L in Buffer 3 (containing protease inhibitor cocktail).
2. Add 1 μ g of antibody and incubate overnight at 4 °C on a rotating wheel. Non-specific [20] and control (e.g., RNA Polymerase II, Histone H3) antibodies must be included in the procedure to measure background levels and monitor the success of the ChIP experiment (*see* **Note 6**).
3. Add 100 μ L of blocked protein A/G beads (prepared in Subheading 3.3.2) in each sample and incubate 2 h at RT on a rotating wheel (*see* **Note 7**).
4. Centrifuge for 5 min at 1,000 $\times g$ at RT and discard supernatant.

3.3.4 Washes

1. Wash the beads twice with 1 mL of cold Low-Salt Buffer: incubate 5 min on a rotating wheel, centrifuge for 1 min at 4,000 $\times g$ and carefully aspirate the supernatant.
2. Wash twice with 1 mL of cold High-Salt Buffer using the same procedure.
3. Wash once with 1 mL of cold Li Buffer using the same procedure.
4. Wash once with 1 mL of TE Buffer using the same procedure.

3.4 DNA Purification

1. Elute DNA–protein complexes by incubating beads in 100 μ L of Elution buffer in a thermomixer (20 min, 65 °C, 1,400 rpm).
2. Centrifuge for 1 min at 4,000 $\times g$ (RT) and transfer the supernatant containing the eluted chromatin in a fresh tube.
3. Repeat **steps 1 and 2** and combine the supernatants (Final volume is 200 μ L).

4. INPUT chromatin is included in the sample processing from here. Add 200 μ L of Elution Buffer and proceed as for the other samples for the next steps.
5. To reverse crosslinking, add 0.2 M NaCl and incubate overnight in a thermomixer (65 °C, 750 rpm).
6. Add 2 μ L of RNase A (0.5 mg/mL) to each sample and incubate 40 min at 37 °C.
7. Add 5 μ L of Proteinase K (20 mg/mL) to each sample and incubate 2–3 h at 55 °C.
8. Purify DNA using phenol/chloroform extraction: add 1 volume of phenol/chloroform solution and transfer into a Phase Lock Gel Light tube. Centrifuge at 4 °C for 5 min at 10,000 $\times g$. Keep the upper phase and precipitate the DNA by adding 3 volumes of 100 % ethanol, sodium acetate 0.3 M, and 5 μ g of glycogen. After 20 min at –80 °C (or overnight at –20 °C), centrifuge at 4 °C for 30 min at 8,000 $\times g$ and carefully discard the supernatant. Finally, wash the pellet with 3 volumes of 70 % ethanol, dry out the pellet and suspend it in 50 μ L of nuclease free water (*see* **Note 8**).
9. Use 2 μ L of purified DNA for PCR/qPCR analysis with standard protocols. Primer design and reactions are the same as for regular PCR/qPCR assays from genomic DNA. In qPCR experiment, results are expressed as percentage of input, which represent the enrichment of the DNA sequence after chromatin IP in regards to the total amount present in your INPUT DNA (Fig. 2).

3.5 Bisulfite Conversion

In this step, the immunoprecipitated DNA is bisulfite converted. As the isolated DNA fragments are relatively small and quantity is limited, regular bisulfite conversion methods are inefficient (*see* **Note 9**).

1. Prepare a serial dilution of DNA with known concentration ranging from 100 ng to 0.025 ng.
2. Prepare the CT Conversion Reagent: add 900 μ L water, 50 μ L of M-Dissolving Buffer, and 300 μ L of M-Dilution Buffer to one tube of CT Conversion Reagent. Mix for 10 min.
3. Add 130 μ L of the prepared CT Conversion Reagent to 20 μ L of sample and standards. Mix by vortexing.
4. Incubate using the following temperature profile: 98 °C for 10 min, 64 °C for 2.5 h, then hold at 4 °C.
5. Add 600 μ L of M-Binding Buffer to a Zymo-Spin IC Column, then add the sample. Close the cap and mix by inverting several times.
6. Centrifuge at full speed (>10,000 $\times g$) for 30 s. Discard the flow-through.

7. Add 100 μL of M-Wash Buffer to the column, spin for 30 s.
8. Add 200 μL of M-Desulphonation Buffer to the column and wait for 15–20 min.
9. Spin at full speed for 30 s.
10. Add 200 μL of M-Wash Buffer to the column, spin for 30 s.
11. Repeat the wash step one more time.
12. Add 10 μL of the M-Elution Buffer directly to the column matrix.
13. Place into a 1.5 mL tube. Spin briefly to elute the DNA.

3.6 Quantification of Amplifiable DNA

The protocol assesses the quantity and integrity of DNA recovered from the bisulfite conversion amplifying an *ALU* repeat consensus sequence in a bisulfite dependent, but methylation-independent manner. The protocol described in this paragraph is based on a publication describing quality control experiments for DNA methylation analysis by MethyLight [19], the probe sequence has been modified to incorporate LNAs instead of the proprietary MGB—quencher to ensure specificity.

1. Prepare a serial dilution of (commercial) bisulfite-treated DNA with a known concentration ranging from 10 ng/ μL to 0.01 ng/ μL .
2. Add 1 $\times$ of probe MasterMix, 0.1 μM probe, 0.3 μM of forward and reverse primer and 3 μL of bisulfite converted DNA (sample or standard) to a final volume of 20 μL .
3. PCR cycling conditions includes an initial denaturation step for 5 min at 95 °C, followed by 55 cycles of 10 s denaturation at 95 °C, 45 s annealing and probe hydrolysis at 60 °C, and 1 s elongation at 72 °C. Acquire fluorescence at the end of the hydrolysis step in each cycle.
4. Determine the amount of amplifiable bisulfite-converted DNA from the C_p -values generated by the instrument's software.

3.7 Design and Optimization of a Pyrosequencing Assay

The assay design is probably the most critical step for Pyrosequencing-based DNA methylation analysis. Great attention should therefore be paid to the experimental design as this will crucially influence the successful outcome of the assay. Many standard software tools developed for conventional PCR cannot handle primer design on bisulfite-converted DNA due to the lower complexity of bisulfite-treated DNA. However, some commercial and freely available software have been especially designed for this purpose such as the CpGWARE primer design software (Chemicon International), MethPrimer [21], Bisearch [22], or MethylPrimer Express (Life Technologies).

3.7.1 *Design of Amplification Primers*

1. As the two strands of genomic DNA are no longer complementary after bisulfite treatment, both strands can be analyzed for possible amplification products. MethPrimer takes only the forward strand into account and the reverse strand has to be created manually or using one of a variety of software tools.
2. Primers should be approximately 30 nucleotides in length and the optimum size of an amplification product is ~250 bp, although we successfully analyzed DNA methylation in PCR products up to 350 bp.
3. Place primers in a region containing four or more cytosines that have been converted during bisulfite treatment to ensure that they are only complementary to completely converted DNA as the chemical treatment is rarely complete.
4. Primers should preferentially contain no CpG positions. If it cannot be avoided the maximum number of CpG positions covered should be restricted to one, and this CpG position should not be included in the last five bases from the 3' terminus to avoid preferential amplification.
5. To ensure specificity palindromes within primers and complementary sequences between primers as well as degenerated bases, inosine should be avoided.
6. The generated amplification product should be verified for the presence of potential polymorphic positions such as SNPs underlying the annealing sites for the amplification primers. We strongly recommend the redesign of amplification primers annealing to potentially polymorphic sites (even at the 5' terminus).

3.7.2 *Design of Pyrosequencing Primers*

1. Identify sequence stretches in the amplification product where a sequencing primer can be positioned and at least the last five bases from the 3' terminus do not overlap with any other potentially variable position including CpGs and SNPs that are retained after bisulfite treatment.
2. Verify the last four or five bases from the 3' terminus to be unique in the amplification product by using, for example, the MS Word find tool. As few as four consecutive nucleotides complementary to a nontargeted stretch of sequence in the amplification product might add to background signal confounding precise quantification.
3. Check successfully designed primers for primer dimers and possible hairpin structures.

4. As read-lengths of up to 120 bases can be achieved with the PyroGold SQA kit, primers can be positioned in nonpolymorphic regions next to the variable region using part of the nucleotide dispensations to approach the region of interest.
5. At least one cytosine not followed by a guanine should be included in the dispensation order to control for complete bisulfite conversion.
6. The direction of the Pyrosequencing primer defines which of the amplification primers needs to be biotinylated. This primer should be checked carefully not to form any hairpin structure.
7. If several Pyrosequencing primers are required to cover the region of interest, analysis of a few CpG positions by more than one Pyrosequencing primer improves confidence into the acquired results and helps to detect potential technical artefacts.

3.7.3 Optimization of the Pyrosequencing Assay

1. Determine the optimal annealing temperature for each primer pair. This step can be performed on a gradient thermocycler using 25 ng of bisulfite-treated genomic DNA as template.
2. Typical PCR reaction conditions are 1× HotStar Taq buffer supplemented with 1.6 mM MgCl₂, 0.1 mM of each dNTP, 2.0 U of HotStar Taq polymerase and 200 nM of forward and reverse primers in a total volume of 25 µL.
3. Typical PCR cycling conditions include an initial denaturation step performed 15 min at 95 °C, followed by 50 cycles of: 30 s denaturation at 95 °C, 15 s of a gradient annealing ranging from 50 to 70 °C and 15 s elongation at 72 °C. Final extension is performed 5 min at 72 °C and the cooling is performed at 4 °C.
4. Deposit 10 µL of PCR product supplemented with 1× loading dye as well as a DNA ladder sample on a 2 % agarose gel in TBE 1× including a fluorescent intercalating dye (e.g., ethidium bromide). Perform a horizontal electrophoresis in TBE 1× at 5 V/cm for 20–30 min. Specificity of the amplifications is assessed by evaluation of the length of the PCR product and the highest temperature allowing a strong specific amplification product is selected as the optimal annealing temperature.

3.8 PCR Amplification

A strong and specific PCR product obtained by amplification with one of the two primers biotinylated is required for a successful Pyrosequencing reaction (*see* **Note 10**).

1. Typical PCR reaction conditions are 1× HotStar Taq buffer supplemented with 1.6 mM MgCl₂, 0.1 mM of each dNTP,

2.0 U of HotStar Taq polymerase, and 200 nM of forward and reverse primer (e.g. 5-AGGAGGGTTTATTTTTTGTAAAGG and 5-Biotin-AAATTCTCCTTCCAAATAAATACTTACAA for the amplification of the *RARβ2* promoter [9, 23]) in a total volume of 25 μL.

2. PCR cycling conditions include an initial denaturation step for 15 min at 95 °C, followed by 50 cycles of 30 s denaturation at 95 °C, 15 s at the optimal annealing temperature (61 °C for the *RARβ2* promoter), and 15 s elongation at 72 °C. Final extension is performed 5 min at 72 °C and the cooling is performed at 4 °C.

3.9 Sample Preparation for Pyrosequencing

This step allows the preparation of the PCR products for Pyrosequencing experiments. It includes the purification and the denaturation of the amplification product, which will be rendered single stranded, and the hybridization of the sequencing primer (*see Note 11*).

1. Transfer 10 μL of PCR product into a standard skirted 96-well PCR plate and add 40 μL of binding buffer, 2 μL of Sepharose beads, and 28 μL of water. Seal the plate and incubate for 10 min at room temperature under constant mixing (1,400 rpm), which is necessary for the capture of the biotinylated PCR products by the Sepharose beads.
2. In parallel, prepare the Pyrosequencing plate by diluting 4 pmol of the Pyrosequencing primer (e.g., 5-TTGAGGA TTGGGATGT and 5-AGGGTTTGT TTTGGGT for the analysis of the CpGs in the *RARβ2* promoter [9, 23]) into 12 μL of annealing buffer into the respective wells of the PSQ plate. Different Pyrosequencing primers can be used in different wells of the same plate.
3. Fill the four troughs of the vacuum preparation tool with 100 mL of 70 % ethanol, 100 mL 0.2 M NaOH denaturing solution, 120 mL of washing buffer, and 150 mL of water (*see Note 12*).
4. Turn on the workstation to create a vacuum in the aspiration device (450 mmHg). Immerse the tips in water for several seconds for cleaning. Then, the PCR plate can be removed from the mixer and the binding mix can be aspirated. The beads remain on the filters of the tips.
5. After all binding mix has been aspirated; immerse the tips in ethanol 70 %, NaOH denaturing solution, and washing solution for 5, 5, and 10 s, respectively. Let the vacuum dry the beads for 10 s and turn it off above the PSQ plate.
6. Immerse the filter tips in the annealing mix of the PSQ plate in order to release the beads into the wells.

7. Heat the sequencing plate for 2 min at 80 °C on a thermoplate placed on a heating device for single DNA strand denaturation. Cool the plate at RT for 5 min for sequencing primer annealing.

3.10 Pyrosequencing Reaction

1. Create a new Pyrosequencing run on the Pyrosequencer using the PyroMark CpG software. Indicate the Pyrosequencing assay used in the selected wells and the name and/or conditions of each wells.
2. Fill the tips of the cartridge with the corresponding reagents and enzyme mix and dNTPs avoiding bubble formation. 180 µL of reagents and enzyme mix is sufficient for a full plate and at least 150 µL of each dNTPs should be used. dNTPs are not limited in the kit and can be used in excess.
3. Deposit a sealed Pyrosequencing test plate and the reagents' cartridge in the Pyrosequencer and perform the dispensation test to verify if the dispensing tips are working properly. Droplets should be clearly visible and homogeneous and the tips should be changed if necessary.
4. Remove the test plate and put the cooled prepared Pyrosequencing plate and start the run. The length of a run is proportional to the number of dispensations (1 per min) and a genotyping run should not exceed 30 min.
5. After the end of the sequencing run, perform a dispensation test to verify if some tips have become blocked during the run (*see Note 13*).
6. Analyze the results with the PyroMark CpG software mode, which allows for the accurate quantification of DNA methylation. To validate the Pyrogram® output of the Pyrosequencing reaction, it is necessary to have at least a signal intensity of 50 a.u. from the beginning until the end of the analyzed sequence for a peak corresponding to the incorporation of a single, nonvariable nucleotide.
7. Export the results in .csv format to enable further treatment and analysis with statistical or graphical software such as Excel®.
8. Remove the cartridge and clean the tips with distilled water. Remaining reagents can be stored at 4 °C and reused, while dNTPs should be discarded.

An example of Pyrograms showing the absence of the co-occurrence of DNA methylation patterns with active gene copies (RNA polymerase II) and the presence of DNA methylation in H3K27me3 and H3K9me3 containing regions in comparison to the overall DNA methylation level (INPUT) is shown in Fig. 3.

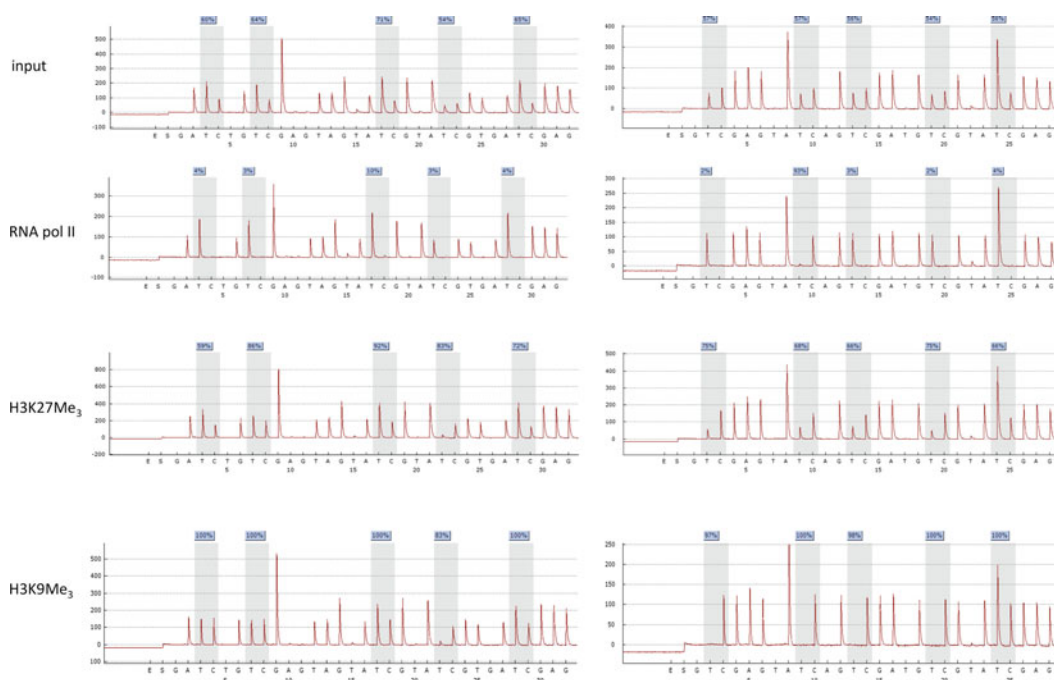


Fig. 3 Pyrograms showing the DNA methylation profile of 10 CpGs of the *RARB2* promoter in the immunoprecipitated DNA using the INPUT or antibodies against RNA polymerase II, H3K27me3, and H3K27me3, respectively. DNA methylation is absent in DNA bound by RNA polymerase II, present at intermediate levels in DNA associated with the histone modification H3K27me3, while DNA associated to H3K9me3 is completely methylated. Analyzed CpG positions are highlighted by grey boxes

4 Notes

1. For most cell lines it is recommended to incubate 10 min at RT with formaldehyde but the time of the crosslinking reaction could require optimization (5–30 min). Over crosslinking may alter antigen accessibility and sonication efficiency. Note that the addition of protein crosslinking reagents such as EGS or DSG prior to formaldehyde treatment could be beneficial if the interaction of the protein of interest with DNA is unstable [24, 25].
2. ChIP experiments can also be performed on native chromatin (N-ChIP), meaning that proteins are not crosslinked to DNA by formaldehyde treatment. This procedure is limited to proteins tightly associated to DNA such as histones and their modifications [26].
3. If using suspension cells, collect 1×10^7 cells with media in a 50 mL tube and perform formaldehyde and glycine treatment as described on a rotating wheel. Pellet cells by centrifugation and wash with cold $1 \times$ PBS. Proceed to **step 6**.

If using thin sections of fresh or frozen tissue, crosslink directly by adding formaldehyde in a sufficient amount of 1× PBS and quench with glycine as described on a rotating wheel. Wash twice with cold 1× PBS. Dissociate mechanically the sample in a sufficient volume of cold Buffer 1, by using a dounce-homogenizer for example, and proceed to **step 7**.

4. Chromatin can also be fragmented by Micrococcal nuclease (MNase), this method is recommended in N-ChIP experiments.
5. A time-course (5, 10, 15, 20, 25, and 30 min) is necessary to determine the optimal sonication conditions for your sample. At each time point, collect 50 µL of shared chromatin and incubate it with RNase A and Proteinase K before loading on a 1.5 % agarose gel. Choose the time point where DNA fragments are in average 200–600 bp in length. Once the sonication time is determined for your sample, keep strictly the same protocol for further experiments as the sonication efficiency depends also on the chromatin concentration. We recommend verifying the size of your shared chromatin at each experiment.
6. Quality and specificity of the antibodies is a critical point in the ChIP procedure. Choose preferentially validated antibodies as “ChIP-grade” from Abcam or “Premium antibodies” from Diagenode, or published antibodies successfully used in ChIP-seq experiments.

It is recommended to titrate the quantity of the antibodies needed to chromatin IP (1–10 µg) as it can differ dependent on the quality/specificity of the antibody and accessibility of the target protein. This step can significantly improve efficiency of the chromatin IP and increase signal to background ratio.

To validate your experiment, include a positive and negative genomic locus in your PCR/qPCR validation step (if using an antibody against a transcription factor, check enrichment at a known promoter binding site and absence of enrichment in a gene body for example).

7. Other protocols use either Protein A or Protein G beads only, but as antibody species and IgG subtypes bind differentially to Protein A or G, we recommend to use resin that combines both types. This allows using the same beads for all antibodies and avoids intersample variations.
8. Alternatively DNA can be purified with a commercial kit such as the “PCR purification kit” from Qiagen, following manufacturer’s instructions. Usually, phenol/chloroform purification yields improved DNA recovery and quality.
9. As the isolated DNA fragments are relatively small and the amount of recovered DNA is limited (normally in the tens of ng after ChIP), bisulfite conversion kits containing a carrier DNA and adapted to fragmented DNA such as DNA extracted

from formalin-fixed paraffin-embedded samples should be used. We compared recovery rates using a number of commercial kits, and the specified kit in this protocol yielded, in reproducible manner, the highest rates. We therefore do not recommend using another kit without extensive validation.

10. Due to the low amount of DNA available for DNA amplification, which could lead to variation in the quantitative estimation of the DNA methylation levels at the analyzed CpG positions [7], PCR amplifications should be performed at least in triplicate.
11. Due to the low amount of DNA available after ChIP and bisulfite conversion the PCR product without a Pyrosequencing primer, the negative (no template) control of the PCR amplification and the Pyrosequencing primer without any PCR product should be processed in parallel for each Pyrosequencing assay to ensure the absence of contaminating amplification product or signals due to primer interactions.
12. A greater volume of washing buffer should be put in the troughs compared to the troughs containing ethanol and NaOH to assure the complete removal of any traces of NaOH, which might inhibit the downstream Pyrosequencing reactions.
13. If the tips were blocked or if no signal was present because of exhausted reagents, the PCR product used for the Pyrosequencing reaction can be resuspended in 20 μ L of binding buffer, and repurified and analyzed by the protocol described in Subheadings 3.9 and 3.10 as previously published [27].

Acknowledgments

This work was supported by ATIP CNRS and Equipe d'Excellence Région Midi-Pyrénées grant to PBA. CM was recipient of an MNERT fellowship.

References

1. Shanmuganathan R, Basheer NB, Amirthalingam L et al (2013) Conventional and nanotechniques for DNA methylation profiling. *J Mol Diagn* 15:17–26
2. Suzuki M, Grealley JM (2013) Genome-wide DNA methylation analysis using massively parallel sequencing technologies. *Semin Hematol* 50:70–77
3. Mensaert K, Denil S, Trooskens G et al (2014) Next-generation technologies and data analytical approaches for epigenomics. *Environ Mol Mutagen* 55:155–170
4. Furey TS (2012) ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat Rev Genet* 13:840–852
5. Tost J, Dunker J, Gut IG (2003) Analysis and quantification of multiple methylation variable positions in CpG islands by Pyrosequencing. *Biotechniques* 35:152–156

6. Tost J, Gut IG (2007) DNA methylation analysis by pyrosequencing. *Nat Protoc* 2: 2265–2275
7. Dupont JM, Tost J, Jammes H et al (2004) De novo quantitative bisulfite sequencing using the pyrosequencing technology. *Anal Biochem* 333:119–127
8. Ahmadian A, Ehn M, Hober S (2006) Pyrosequencing: history, biochemistry and future. *Clin Chim Acta* 363:83–94
9. Moison C, Assemat F, Daunay A et al (2014) Synergistic chromatin repression of the tumor suppressor gene RARB in human prostate cancers. *Epigenetics* 9:477–482
10. Carey MF, Peterson CL, Smale ST (2009) Chromatin immunoprecipitation (ChIP). *Cold Spring Harb Protoc.* 2009:pdb prot5279
11. Kuo MH, Allis CD (1999) In vivo cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. *Methods* 19:425–433
12. Orlando V (2000) Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. *Trends Biochem Sci* 25:99–104
13. Johnson KD, Bresnick EH (2002) Dissecting long-range transcriptional mechanisms by chromatin immunoprecipitation. *Methods* 26:27–36
14. Ren B, Dynlacht BD (2004) Use of chromatin immunoprecipitation assays in genome-wide location analysis of mammalian transcription factors. *Methods Enzymol* 376:304–315
15. Bernstein BE, Humphrey EL, Liu CL et al (2004) The use of chromatin immunoprecipitation assays in genome-wide analyses of histone modifications. *Methods Enzymol* 376:349–360
16. Bernstein BE, Mikkelsen TS, Xie X et al (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125:315–326
17. Brinkman AB, Pennings SW, Braliou GG et al (2007) DNA methylation immediately adjacent to active histone marking does not silence transcription. *Nucleic Acids Res* 35:801–811
18. Statham AL, Robinson MD, Song JZ et al (2012) Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. *Genome Res* 22:1120–1127
19. Campan M, Weisenberger DJ, Trinh B et al (2009) MethyLight. *Methods Mol Biol* 507:325–337
20. Kramer A, Mailand N, Lukas C et al (2004) Centrosome-associated Chk1 prevents premature activation of cyclin-B-Cdk1 kinase. *Nat Cell Biol* 6:884–891
21. Li LC, Dahiya R (2002) MethPrimer: designing primers for methylation PCRs. *Bioinformatics* 18:1427–1431
22. Aranyi T, Varadi A, Simon I et al (2006) The BiSearch web server. *BMC Bioinformatics* 7:431
23. Moison C, Senamaud-Beaufort C, Fourriere L et al (2013) DNA methylation associated with polycomb repression in retinoic acid receptor beta silencing. *FASEB J* 27:1468–1478
24. Kurdistani SK, Grunstein M (2003) In vivo protein-protein and protein-DNA crosslinking for genomewide binding microarray. *Methods* 31:90–95
25. Nowak DE, Tian B, Brasier AR (2005) Two-step cross-linking method for identification of NF-kappaB gene network by chromatin immunoprecipitation. *Biotechniques* 39:715–725
26. O'Neill LP, Turner BM (2003) Immunoprecipitation of native chromatin: NChIP. *Methods* 31:76–82
27. Tost J, El abdalaoui H, Gut IG (2006) Serial pyrosequencing for quantitative DNA methylation analysis. *Biotechniques* 40:721–726

Free ebooks ==> www.Ebook777.com

Part IV

Bacterial Typing and Identification

Free ebooks ==> www.Ebook777.com

Chapter 23

Multiplex Pyrosequencing®: Simultaneous Genotyping Based on SNPs from Distant Genomic Regions

Piotr Wojciech Dabrowski, Kati Bourquain, and Andreas Nitsche

Abstract

Pyrosequencing® is a technique that allows the quick sequencing of short stretches of DNA, and results can be inspected in real time as they are generated. As such, it is a highly useful tool for the typing of organisms, based on single-nucleotide polymorphisms (SNPs). However, if a single polymorphism is not sufficient for typing and several distant genomic regions need to be examined, multiple Pyrosequencing reactions have to be performed for each organism. This strongly increases both workload and reagent costs. Alternatively, multiplex Pyrosequencing can be performed, in which the multiple sequencing reactions for all analyzed genomic regions are performed in a single reaction. However, when using this method, special care has to be taken while designing the assay and analyzing the results.

Here we present a detailed protocol using our newly developed software mPSQed for assay design and MultiPSQ for data analysis.

Key words Pyrosequencing®, Multiplex Pyrosequencing®, Genotyping, Data analysis, Assay design

1 Introduction

1.1 Pyrosequencing

Pyrosequencing is a method of sequencing by synthesis which was first described in 1996 [1]. The principle is visualized and described in Fig. 1 and described in Chapters 1 and 2.

In short, the complementary strand to a single-stranded DNA template is synthesized by a polymerase. However, instead of providing a mix of all four dNTPs, only one of the possible dNTPs is provided at a time. The enzymes sulfurylase and luciferase are used to convert the pyrophosphate, which is released when a dNTP is incorporated, into a light signal which in turn is recorded by a camera. The sequence of the template can then be reconstructed from the intensities of the light signals recorded when the specific dNTPs are added. A “Pyrogram®” is the plot which shows the signal intensity generated over time and in function of the time at which the different dNTPs were added to the reaction. Since the light signal occurs at the same time as the dNTP is added to the

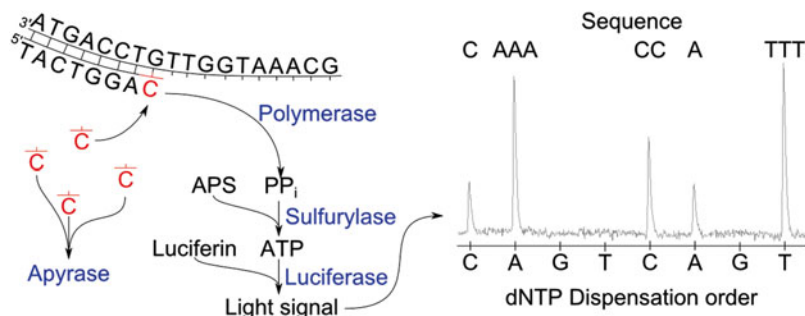


Fig. 1 The principle of Pyrosequencing. In each cycle of the reaction, one of the four possible dNTPs is added to the reaction. Depending on the sequence of the template, none, one, or several of the available dNTP molecules can be incorporated at the 3' end of the newly synthesized strand. If a dNTP molecule is incorporated, pyrophosphate is split off during the reaction, which is used by ATP sulfurylase to produce ATP from ADP. This in turn provides energy for luciferase to oxidize luciferin—a reaction leading to the emission of a photon. The resulting light signal can be detected using a camera. If more than one dNTP molecule of the same species can be incorporated, proportionally more pyrophosphate is generated, leading to a proportionally stronger light signal. Finally, before the next addition step of a new dNTP species is started, the remaining, non-incorporated dNTP molecules are degraded by apyrase. Through correlation of the order of dNTPs which have been added to the reaction mix and the order and intensity of the light signals the sequence of the template can be reconstructed

reaction mix, the incorporation of each successive dNTP can be visualized in real time. This is advantageous in genotyping settings where the assay is usually designed to use the first few bases of the sequence to determine the genotype of the organism in the sample. Instead of having to wait for the whole sequencing reaction to finish, as would be the case with Sanger sequencing, the user can see the signature bases immediately and abort the run as soon as the results are conclusive, thus saving time.

1.2 Multiplex Pyrosequencing

While Pyrosequencing offers the advantage of being fast and presenting the results in real time, it offers only a limited read length. If complex genotyping applications require the use of several SNPs which are located far away from each other on the genome, a separate PCR product must be generated and sequenced for each of the SNPs. This increases both the workload and the cost of chemicals. One way to overcome this limitation is the use of multiplex Pyrosequencing [2–6]. In multiplex Pyrosequencing, the products of several PCR reactions, each of which creates a template containing discriminating SNPs from a different genomic region, are sequenced in the same well. This means that when a dNTP is provided to the polymerase during a cycle, a light signal is generated if it can be incorporated into any of the available templates; several Pyrosequencing reactions take place at the same time in the same reaction vessel, and the signals generated by all the individual reactions overlap. The resulting Pyrogram thus cannot be used to reconstruct the sequence of the templates directly. However, if the sequences of the templates in the reaction are known, it is possible

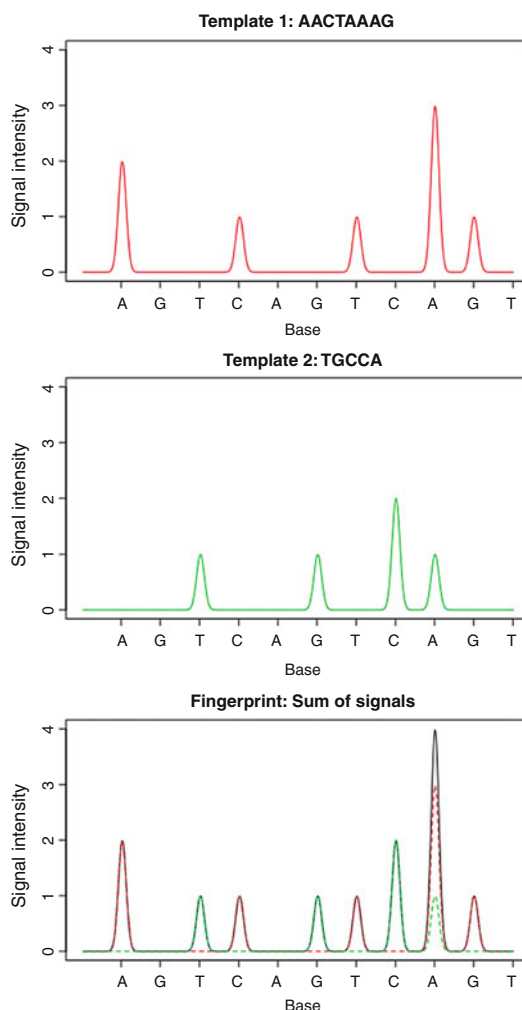


Fig. 2 The principle of multiplex Pyrosequencing. As in Pyrosequencing, light signals are generated when dNTPs are incorporated into the template. However, since several templates (two in this example) are available, dNTP incorporation into any of the templates generates a light signal and the signals overlap, creating a fingerprint Pyrogram. Given only the fingerprint, it is not possible to reconstruct the template sequences because one cannot know which template was responsible for which signal

to pre-calculate the Pyrogram which would result from the multiplex Pyrosequencing reaction (Fig. 2).

Genotyping applications usually target specific SNPs in well-known sequences. Assuming that one is looking for known genotypes, and not trying to identify new ones, it is safe to say that all possible (or at least all expected) template sequences are known. Thus it is possible to create a list of the expected template sequences for each genotype and to calculate the resulting fingerprints. Once the expected fingerprint for each genotype is known, multiplex

Pyrosequencing can be performed. In this case it is no longer necessary to try to recreate the exact sequences of the multiple templates from the resulting Pyrogram—it is sufficient to match the Pyrogram to one of the previously calculated expected fingerprints in order to determine the genotype of the sample.

2 Materials

2.1 Assay Design

1. Alignment of the genomes of the different genotypes which are to be detected.
2. A tool for designing multiplex Pyrosequencing assays (e.g., mPSQed [7]).
3. A tool for assessing PCR primers (e.g., <http://www.thermo-scientificbio.com/webtools/multipleprimer/>).

2.2 Sample Preparation

1. QIAamp DNA Mini Kit (Qiagen).

2.3 PCR with Biotinylated Primers

1. One unmodified primer.
2. One biotinylated primer for PCR amplification.
3. PCR buffer (Thermo Fisher Scientific).
4. Platinum® Taq DNA polymerase (Thermo Fisher Scientific).
5. dNTPs.
6. MgCl₂.
7. 96-Well PCR plates, skirted.
8. Thermocycler (e.g., Mastercycler (Eppendorf)).

2.4 Gel Electrophoresis and Quantification

1. Agarose.
2. Ethidium bromide 0.5 µg/mL.
3. Loading buffer.
4. Low mass DNA ladder (Thermo Fisher Scientific).
5. 1× TAE buffer.
6. E.A.S.Y RH-3 (software by Herolab, Wiesloch, Germany).

2.5 Multiplex Pyrosequencing®

1. Streptavidin Sepharose HP beads (GE Healthcare).
2. PyroMark® Q96 Vacuum Prep Workstation (Qiagen).
3. PyroMark Binding Buffer (Qiagen).
4. 70 % ethanol.
5. Denaturation solution: 0.2 M NaOH.
6. PyroMark Wash Buffer (Qiagen).
7. PyroMark Annealing Buffer (Qiagen).
8. Thermowell sealing tape.

9. Thermomixer or similar (room temperature).
10. Heating device, for example thermoblock.
11. 96-Well plate for sample preparation.
12. Primers for multiplex Pyrosequencing.
13. PyroMark Q96 Plate Low (Qiagen).
14. PyroMark Q96 Cartridge (Qiagen).
15. PyroMark Gold Q96 Reagents (Qiagen).
16. PyroMark ID (Qiagen).

2.6 Analysis of Multiplex Pyrograms

1. PyroMark software.
2. Multiplex Pyrosequencing® assay analysis software (e.g., multiPSQ [2] or AdvISER-PYRO [8]).

3 Methods

3.1 Assay Design

In order to design a multiplex Pyrosequencing assay, it is necessary to first identify polymorphisms, which allow discrimination between genotypes. For this, an alignment of the target sequences in the FASTA format is required. While an in-depth description of the creation of such an alignment would be outside the scope of this chapter, a general approach might be to:

1. Determine a region of interest, e.g., via literature search.
2. Extract the sequence of this region from one known genome sequence.
3. Perform a BLAST search of this sequence (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>).
4. Download the matching sequences and align them (*see Notes 1 and 2*). When using many sequences, consider using an alternative to ClustalW [9] (such as mafft [10]), since ClustalW is very slow when computing large multiple alignments.

Once the alignment has been generated, the SNPs which can be used for genotyping have to be determined. This can be done by using a program such as mPSQed [7], which will be used in this example.

In mPSQed, select the option “File→Load file” to open the alignment. The names of the sequences will be displayed on the left side and the sequences themselves on the right side. To zoom in or out, press + or -. For the identification of SNPs which can be used for genotyping, sequences must be assigned to different genotypes. This can be done by selecting sequence names on the left side, then by right clicking on one of the selected names, and selecting the option “Set as group” (which allows the definition of a new genotype) or “Add sequence to group” (which moves the selected



Fig. 3 An alignment loaded into mPSQed (*left*) and the view once groups have been assigned and collapsed and SNPs have been identified (*right*): The *green/orange graphs* show the conservation within each group of sequences, and the discriminating SNPs are marked as *red columns*

sequences to a previously defined group). Once all sequences are assigned to genotypes, SNPs can be identified by selecting the option “Analyze→Find all SNPs” (Fig. 3). All SNPs that can be used to differentiate between genotypes are highlighted. Then the alignment has to be inspected in order to find at least one unique SNP for each genotype. Preferably, several of the SNPs should be located sufficiently near each other to allow for amplification and sequencing with a single pair of primers—this reduces the number of different reactions which have to be performed together and may interfere with one another.

Once the SNP positions have been identified, primers for both PCR and Pyrosequencing need to be designed. This can also be done in mPSQed: select the bases the primer is supposed to bind to, right-click on the selected region, and select the option “Add primer→Forward” or “Add primer→Reverse”. If both the forward and reverse primers are added to the same sequence, mPSQed will display a line connecting these primers which shows the length of the product and will highlight the line in red if the melting temperatures of the primers are further apart than 10 °C (this can be changed under “Edit→Settings”). The melting temperature of the primers is displayed to their left and the length is displayed to their right. Primers can be moved by dragging them to the sides and resized by grabbing the edges and dragging these.

Once PCR primers have been defined, the Pyrosequencing primers need to be designed (*see Note 3*). These must lie within the parts which will be amplified by the PCR primers and should be as close as possible to the SNPs which are to be used for genotyping (*see Notes 4 and 5*). Pyrosequencing primers are first added as normal primers, and are then converted to Pyrosequencing primers by right-clicking on them and selecting “Is Pyrosequencing primer”. An example of an alignment with a primer pair and a Pyrosequencing primer is given in Fig. 4.

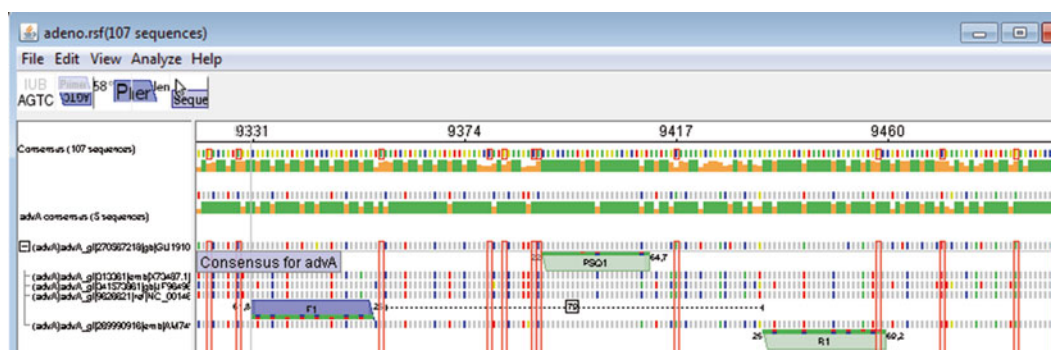


Fig. 4 Alignment in mPSQed with two PCR primers and a Pyrosequencing primer

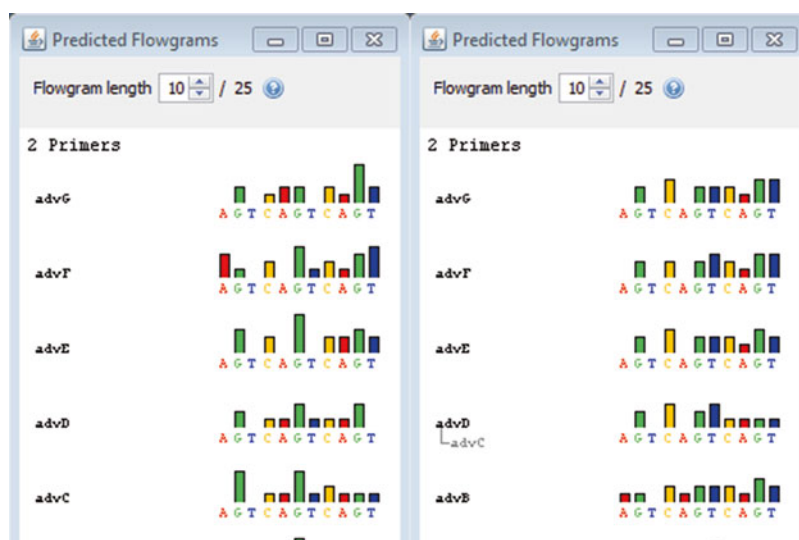


Fig. 5 View of predicted flowgrams. On the *left side*, each sequence group creates a unique fingerprint. On the *right side*, the groups advD and advC create the same fingerprint, meaning that they cannot be discerned using the given assay

When placing the Pyrosequencing primers, it is critical to ensure that the resulting fingerprints will be unique for each of the genotypes. This can be checked by selecting the option “Analyze→Show predicted flowgrams”. This shows the fingerprints, which are expected to be generated for each of the defined genotypes using the defined Pyrosequencing primers (Fig. 5). Should the predicted flowgrams not be unique, either the flowgram length can be increased (using the option on top of the window, *see Note 3*) or the Pyrosequencing primers can be moved by a few bases. The window that shows the predicted flowgrams is updated automatically.

Once the primers lead to predicted flowgrams suited for genotyping, an assay definition file—which is necessary for the analysis of signals from the Pyrosequencing machine—can be exported using the option “File→Export assay definition”. It is advisable to check the primers for unwanted interactions such as primer–primer dimers or secondary structures using a specialized tool (such as <http://www.thermoscientificbio.com/webtools/multipleprimer>, which is able to analyze a large number of primers simultaneously) before proceeding to setting up the assays in the lab (*see* **Note 6**).

3.2 Sample Preparation

1. Extract the (viral) DNA using commercial kits.

3.3 PCR with Biotinylated Primers

1. Prepare the PCR reaction mixture. Each 25 μL reaction contains 3 μL of DNA, 1 $\times$ PCR buffer, 4 mM MgCl_2 , 200 nM of each dNTP, 800 nM of each primer, one of them being biotinylated, and 1 U Platinum[®] Taq polymerase (*see* **Note 6**).
2. Amplify the targeted region using a PCR program consisting of an initial denaturation step of 5 min at 95 $^{\circ}\text{C}$, followed by 45 cycles of 15 s at 95 $^{\circ}\text{C}$, 30 s at 54 $^{\circ}\text{C}$, and 30 s at 72 $^{\circ}\text{C}$, with a final extension of 5 min at 72 $^{\circ}\text{C}$.
3. PCR products can be stored at +4 $^{\circ}\text{C}$ for several days or at –20 $^{\circ}\text{C}$ for several months.

3.4 Gel Electrophoresis and Quantification

1. Deposit 2 μL of PCR product with 1 μL of 6 $\times$ loading buffer on a 1.5 % agarose gel (*see* **Note 7**).
2. Load 2 μL of low-molecular-weight DNA ladder in parallel.
3. Electrophoresis is performed with 1 $\times$ TAE buffer at 6–8 V/cm for 30 min.
4. Resulting bands are quantified using the E.A.S.Y RH-3 software (Herolab).

3.5 Sample Preparation for Pyrosequencing Analysis

1. Transfer 20 μL of the PCR product into a new PCR plate, add 40 μL of binding buffer and 3 μL of Sepharose beads, and complete to 80 μL with water. Cover the plate with a sealing tape and incubate for 5 min at room temperature and for another 5 min under constant shaking (1,400 rpm) to make sure that the beads do not sediment.
2. During the incubation, prepare the Pyrosequencing plate by diluting 10 pmol of each sequencing primer into 40 μL of annealing buffer into the respective wells of the PSQ plate. Multiple PSQ primers are used in the same well.
3. Fill the troughs of the vacuum workstation with 70 % ethanol, NaOH (0.2 M), washing buffer, and water, respectively.
4. Turn on the workstation and create a vacuum in the aspiration device. Remove the PCR plate from the mixer and aspirate the binding mix.

5. Immerse the tips of the filters for 5 s each in the successive baths of ethanol, denaturation solution, and washing buffer.
6. Place the tips over the Pyrosequencing plate, turn off the vacuum, and immerse the tips of the filters in the annealing mix. Shake gently to release the beads into the wells.
7. Incubate the sequencing plate for 2 min at 80 °C on the thermoplate placed in a heating device. Allow the plate to cool down to room temperature.

3.6 Pyrosequencing Reaction

1. During cooling of the sequencing plate, program the sequencing run on the Pyrosequencer. The default dispensation order when using assays designed with mPSQed is AGTC.
2. The software calculates the quantity of reagents necessary to perform the run. Dispense the reagents and enzyme mix in the appropriate tips in the cartridge. Be careful not to create air bubbles.
3. Deposit the sequencing plate and the reagents cartridge into the Pyrosequencer.
4. Start the Pyrosequencing run.
5. After the run is finished, remove the cartridge and clean the tips well with pure water.

3.7 Analysis of Multiplex Pyrograms

Once the Pyrosequencing reaction is finished, the data needs to be exported from the sequencer in order to analyze it. This can be done by starting the Pyrosequencing Data Exchange Tool on the control PC. In the Exchange Tool, log into the export database (the username and password can be obtained from the machine operator; the default username and password are both “user”). Then click on “Export”, select the appropriate run in the left window, click on “browse”, and select a file for saving the data. The file name has to end with the suffix “.xml”; otherwise it will not be correctly loaded into the analysis tool.

While there are different analysis tools (e.g., AdvISER-PYRO [8]), in this chapter the analysis using MultiPSQ [2] is described. The analysis consists of the following steps:

1. Start the program “Pyrosequencer.exe”.
2. Select the option “File→Open Assay Definition File”.
3. Navigate to and select the assay definition file saved during assay creation (see the section “Assay Design”).
4. Select the option “File→Open Flowgram File”.
5. Navigate to and select the xml file exported from the control PC.
6. Click on “Analyze all”.
7. To see the result for each well, click on the well in the top right part of the window. The flowgram and the classification will be displayed (see Fig. 6) (see **Notes 8** and **9**).

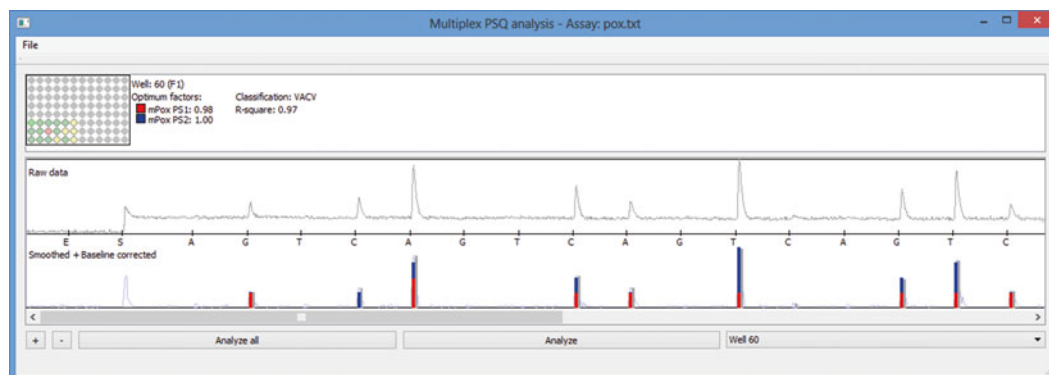


Fig. 6 View of a Pyrosequencing run after classification. The plate is shown in the *top left* corner. Unused wells are *grey*, *green* indicates a valid classification, *yellow* a negative control, and *red* an unrecognized fingerprint. Next to the plate, the classification for the selected well is displayed. Below, the Pyrogram and the recognized fingerprint, including the contribution of each primer to the signal, are shown

4 Notes

1. It is important to include all known sequences from the organism under investigation in the assay design. Failing to do so is likely to lead to misclassifications.
2. If the downloaded sequences seem not to align sufficiently well, check their direction. Some of the downloaded sequences may be from the positive and some from the negative strand. In such a case, most alignment algorithms will produce no meaningful result. It is thus necessary to ensure that all sequences have the same direction, e.g., by reverse-complementing all sequences from the negative strand.
3. Not every SNP which differentiates between genotypes can be used: Sometimes, it may be in a region where it is hard to design primers or the signal may interfere with that from other SNPs. If in doubt, try to find a different SNP.
4. The quality of the Pyrosequencing signal is best at the beginning of a sequence. Try to design assays so that as few bases as possible have to be sequenced by putting Pyrosequencing primers directly before the relevant SNPs.
5. Make sure that the Pyrosequencing primer and the biotinylated PCR primer are on different strands—during the Pyrosequencing reaction, only the strand containing the biotinylated primer is present, so the Pyrosequencing primer needs to be able to bind to that.
6. Even if *in silico* analysis indicates no incompatibilities between the primers, it is best to establish all PCR and Pyrosequencing reactions, which are to be used in the multiplex assay, separately before multiplexing.

7. While this step is important when establishing the method, it can be left out during assay routine application in order to decrease the time needed.
8. If one of the primers seems not to contribute to any of the signals (all bars have only one color), one of the PCRs may be too inefficient. Perform the PCRs individually, quantify each of the products individually, and, based on the quantification, use the same amount of each PCR product in the multiplex Pyrosequencing. If this changes the results, modify the PCR primers to achieve equal efficiency of all PCR reactions.
9. If a good Pyrogram is generated but the fingerprint is not recognized by multipsq, it is advisable to perform individual Sanger sequencing of the PCR products, since this indicates the presence of a novel mutation.

References

1. Ronaghi M, Karamohamed S, Pettersson B et al (1996) Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem* 242:84–89
2. Dabrowski PW, Schroder K, Nitsche A (2013) MultiPSQ: a software solution for the analysis of diagnostic n-plexed pyrosequencing reactions. *PLoS One* 8:e60055
3. Pourmand N, Elahi E, Davis RW et al (2002) Multiplex Pyrosequencing. *Nucleic Acids Res* 30:e31
4. Patel P, Shiao YH, Fortina P (2007) Multiplex pyrosequencing for DNA variation analysis. *Methods Mol Biol* 373:75–88
5. Gharizadeh B, Oggionni M, Zheng B et al (2005) Type-specific multiple sequencing primers: a novel strategy for reliable and rapid genotyping of human papillomaviruses by pyrosequencing technology. *J Mol Diagn* 7: 198–205
6. Doostzadeh J, Shokralla S, Absalan F et al (2008) High throughput automated allele frequency estimation by pyrosequencing. *PLoS One* 3:e2693
7. Dabrowski PW, Nitsche A (2012) mPSQed: a software for the design of multiplex pyrosequencing assays. *PLoS One* 7:e38140
8. Ambroise J, Piette AS, Delcorps C et al (2013) AdvISER-PYRO: amplicon identification using SparsE representation of PYROsequencing signal. *Bioinformatics* 29:1963–1969
9. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680
10. Katoh K, Misawa K, Kuma K et al (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30:3059–3066

Free ebooks ==> www.Ebook777.com

Chapter 24

Detection of Drug-Resistant *Mycobacterium tuberculosis*

Anna Engström and Pontus Juréen

Abstract

Tuberculosis (TB) remains a global health problem. The increasing prevalence of drug-resistant *Mycobacterium tuberculosis*, the causative agent of TB, demands new measures to combat the situation. Rapid and accurate diagnosis of the pathogen and its drug susceptibility pattern is essential for timely initiation of optimal treatment, and, ultimately, control of the disease. We have developed a molecular method for detection of first- and second-line drug resistance in *M. tuberculosis* by Pyrosequencing®. The method consists of seven Pyrosequencing assays for the detection of mutations in the genes or promoter regions, which are most commonly responsible for resistance to the drugs rifampicin, isoniazid, ethambutol, amikacin, kanamycin, capreomycin, and fluoroquinolones. The method was validated on clinical isolates and it was shown that the sensitivity and specificity of the method were comparable to those of Sanger sequencing. In the protocol in this chapter we describe the steps necessary for setting up and performing Pyrosequencing for *M. tuberculosis*. The first part of the protocol describes the assay development and the second part of the protocol describes utilization of the method.

Key words Pyrosequencing®, *rpoB*, *inhA*, *katG*, *gyrA*, *rrs*, *eis*, Mycobacteria, Resistance

1 Introduction

Drug resistance is a global health concern. Resistant bacteria hamper infection control, and often lead to prolonged illness, increased health care costs, and greater risk of mortality, and ultimately threaten to bring us back to the pre-antibiotic era. Resistant *Mycobacterium tuberculosis* is no exception in this case. Strategies to control tuberculosis are hindered by the global emergence and spread of drug-resistant strains. This sets the importance of diagnostics into a new perspective. It has become evident that it is crucial to detect not only *M. tuberculosis*, but also its drug susceptibility pattern. This is particularly important in settings with high prevalence of drug-resistant strains. Traditional culture-based drug susceptibility test methods for *M. tuberculosis* are inherently time and labor intensive; therefore, it is recommended to use molecular diagnostic methods for a more rapid and simplified detection of drug-resistant *M. tuberculosis* [1].

Screening for mutations associated with drug resistance in *M. tuberculosis* can be performed by a number of molecular methods, whereof most are based on PCR in combination with subsequent probe or amplicon hybridization [2–7]. These methods, however, target, sometimes indirectly, only some and not all mutations linked to resistance, and as a consequence clinically relevant mutations may not be detected or misclassified leading to incorrect results. Silent mutations are particularly problematic for probe-based detection methods since they may cause loss of wild-type signal, incorrectly indicating a resistant phenotype. Sequencing, on the other hand, does not discriminate against any type of mutation, no matter if novel or a priori known, missense, or silent. Thus, sequencing has the potential to yield a better sensitivity and specificity than other molecular diagnostic methods. As sequencing is accurate and robust, the methodology has widely been used for characterizing mutations in *M. tuberculosis* [8–13] and it is often considered as the gold standard for resistance detection with molecular methods [14].

Pyrosequencing[®] relies on the detection of pyrophosphate (PPi) release upon nucleotide incorporation, rather than chain termination with ddNTPs and subsequent fragment separation by electrophoresis as in Sanger sequencing. In Pyrosequencing, one of the four nucleotides is added separately to the reaction. If the DNA polymerase incorporates the correct, complementary dNTPs onto the template, PPi is released [15]. The enzyme adenosine 5'-triphosphate (ATP) sulfurylase quantitatively converts PPi to ATP in the presence of adenosine 5' phosphosulfate. This ATP acts as fuel to the luciferase-mediated conversion of luciferin to oxyluciferin which generates visible light in amounts that are proportional to the amount of ATP, i.e., the number of nucleotides incorporated. The light produced in the luciferase-catalyzed reaction is detected by a camera. Unincorporated nucleotides and ATP are degraded by the apyrase, and the reaction can start again with another nucleotide addition.

Pyrosequencing is a robust technique that is easy to perform in a large-scale manner with its 96-well format. The technique holds great flexibility as up to 96 different assays can be performed in parallel in one single run or one assay can analyze 96 samples simultaneously. Several 96-well plates containing samples can easily be prepared and sequenced after each other. The vast majority of the described *M. tuberculosis* resistance-associated mutations are located in relatively short regions within genes or promoter regions which can readily be covered by Pyrosequencing.

We have developed an extensive method for the detection of first- and second-line drug resistance in *M. tuberculosis* [10]. The method consists of seven Pyrosequencing assays for the detection of mutations in the genes or promoter regions which are most commonly responsible for resistance to the drugs rifampicin, isoniazid,

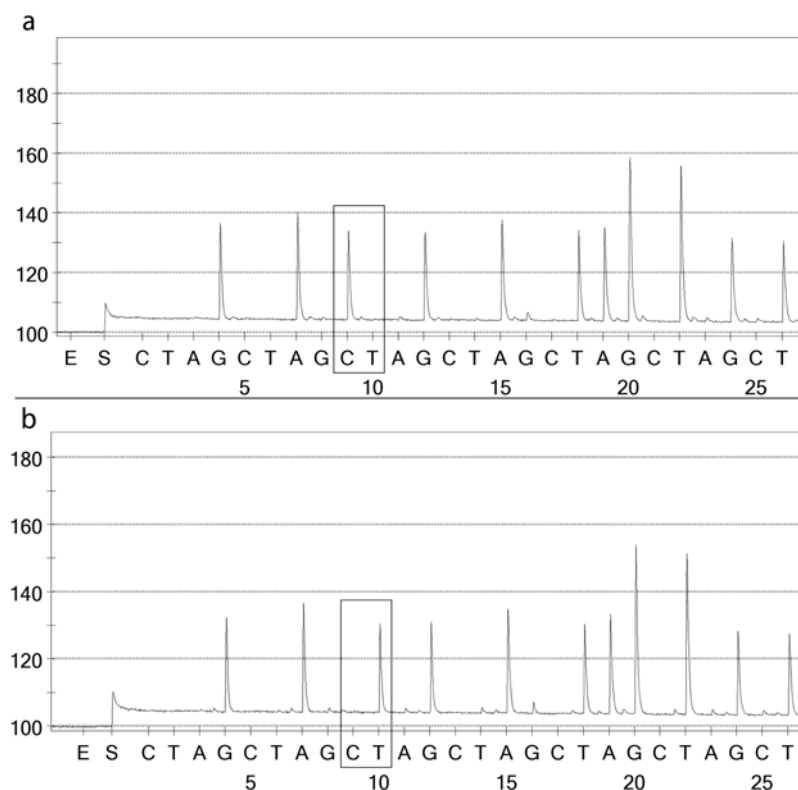


Fig. 1 Pyrograms of the *inhA* promoter region assay. The boxes indicate the nucleotide at position –15 which in wild-type *M. tuberculosis* strains is a C (a) and in the mutant strain a T (b). Nucleotide positions –17 and –8 are also covered by the assay and are in both these cases wild type

ethambutol, amikacin, kanamycin, capreomycin (CAP), and fluoroquinolones. The genes *rpoB*, *katG*, *embB*, *rrs*, and *gyrA*, and the promoter regions of the *mabA-inhA* operon (commonly referred to as *inhA*) and *eis* were covered [10]. Wild-type and mutant variants of the *inhA* promoter region are presented in Fig. 1.

The method is easy to use and robust, and it was shown that, despite that Pyrosequencing renders shorter sequencing reads (typically 30–60 bp) compared to Sanger sequencing, the sensitivity and specificity of the method were similar to those of Sanger sequencing [9]. For the *rrs* gene however, we were able to achieve a read length of 95 bp, facilitated by few possible mutations (at nucleotide positions 1401, 1402, and 1484) and directed dispensation order (Fig. 2).

The limitation of the read length is mainly due to dilution effects in the sequencing reaction, as described in detail in Chapter 1. The read length limits investigation of large genomic regions, e.g., detection of *pncA* gene mutations which are linked to pyrazinamide resistance [16] and *tlyA* mutations which are associated with resistance to CAP [17]. The relevance of *tlyA* mutations has been

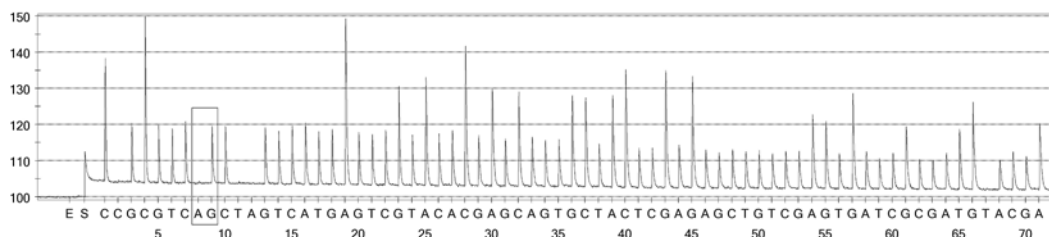


Fig. 2 Pyrogram of the *rrs* assay, which yields a sequence length of 95 bp. The *M. tuberculosis* strain in this example harbors a mutation from A to G at nucleotide position 1401 (*boxed*)

questioned though, as it has been shown that *tlyA* mutations are rarely present in CAP-resistant *M. tuberculosis* of clinical origin and not all mutations in *tlyA* seem to lead to CAP resistance in *M. tuberculosis* [18]. Our method was validated on DNA from cultured clinical isolates [10]; however, Pyrosequencing has also successfully been employed directly on sputum samples [19].

In this chapter we describe the steps necessary for setting up and performing Pyrosequencing for *M. tuberculosis*. The first part of the protocol describes the assay development and the second part of the protocol describes utilization of the method.

2 Materials

2.1 Assay Design

1. Design of PCR primers can be done manually, but preferably by using a software, e.g., the commercial PyroMark Assay Design SW 2.0 (Qiagen GmbH, Hilden, Germany, formerly Pyrosequencing AB) (*see Note 1*).
2. Design of Pyrosequencing primers can be done manually, but preferably by using the commercial PyroMark Assay Design SW 2.0 (Qiagen).

2.2 Sample Preparation

1. 50 mM Tris-HCl, pH 8.0.
2. Water bath (85 °C).
3. Ultrapure water.
4. Chloroform.
5. Screw-cap microcentrifuge tubes (1.5 mL).
6. MiniSpin centrifuge (Eppendorf, Hamburg, Germany).
7. Vortex.
8. NanoDrop spectrophotometer (NanoDrop Technologies Inc, Wilmington, DE).

2.3 PCR Amplification

1. Primers for PCR amplification (assay specific). *See Table 1* for primer sequences.

2. Biotinylated primers for PCR amplification (*see Note 2*).
3. AmpliTaq Gold provided with PCR buffer and MgCl₂ (Thermo Fisher Scientific, formerly Applied Biosystems).
4. dNTPs.
5. Thermocycler: GeneAmp PCR System 9700 (Thermo Fisher Scientific, formerly Applied Biosystems).

All reagents should be stored at -20 °C.

Table 1
Primer sequences

Gene	Primer name	Type of primer ^a	Sequence (5'-3')
<i>rpoB</i>	rpoB.PCR.SGS.Forw2	F	TTTCGATCACACCGCAGACGTT
	rpoB.biotin.Rev2	R	Biotin-GGCACGCTCACGTGACAGAC
	F1	S	GCGATCAAGGAGT
	F7	S	CAGCCAGCTGAGCCAATTCA
	F9	S	ACCAGAACAACCCGCTGTCTG
	F13	S	TGACCCACAAGCGCCGACTG
<i>katG</i>	katG.PCR.F1	F	TGGAGCAGATGGGCTTGG
	katG.PCR.R1	R	Biotin-TCGTAGCCGTACAGGATCTCGA
	katG.seq.S2	S	CGGTAAGGACGCG
<i>inhA</i>	inhA_PSQ_F2	F	AAAGTTCCCGCCGAAATC
	inhA_PSQ_R2	R	Biotin-CTGAACGGGATACGAATGGG
	inhA_AE_S1	S	GCCCGGCCGCGGCGA
<i>embB</i>	FEMBP	F	TCCTGCTCTGGCATGTCAT
	REMBP	R	Biotin-GCCGAACCAGCGGAAATA
	S1EMBP	S	GCTACATCCTGGGC
<i>gyrA</i>	AE.gyrA.PCR.F1	F	TGTTTCGATTCCGGCTTCC
	AE.gyrA.PCR.R1biotin	R	Biotin-CGGGCTTCGGTGTACCTCAT
	GyrA.seq.F1	S	AACTACCACCCGCAC
<i>rrs</i>	FrrsP	F	TATAGGTAAAGCGAATCCTTAAAAGCC
	rrsP-Rbi	R	Biotin-GGCGTTTTTCGTGGTGCTC
	SrrsO1400	S	CGGGCCTTGTACACA
<i>eis</i>	eisPyroF1	F	TTCATCGCGTGATCCTTTGC
	eisPyroR2-bio	R	Biotin-GTGTAGCCCGACCGAGGA
	eisPyro1seq	S	AGACACTGTCGTCGTAAT

^a F forward PCR primer, R, reverse PCR primer, S Pyrosequencing primer

2.4 Sample Preparation for Pyrosequencing Analysis

1. Streptavidin-coated Sepharose High Performance (HP) beads (GE Healthcare, Uppsala, Sweden).
2. PyroMark® Q96 Vacuum Workstation [20] (Qiagen) using corresponding PyroMark Vacuum Prep Filter Probes (Qiagen).
3. PyroMark Q96 Vacuum Prep Troughs (Qiagen).
4. PyroMark Binding Buffer (Qiagen) (*see Note 3*).
5. PyroMark Denaturation Solution (Qiagen).
6. PyroMark Wash Buffer (Qiagen).
7. PyroMark Annealing Buffer (Qiagen).
8. 70 % ethanol.
9. Ultrapure water.
10. ThermoMixer C (Eppendorf) or similar for 96-well plate (room temperature).
11. Heating device, e.g., heating plate or thermoblock.
12. PyroMark Q96 Sample Prep Thermoplate Low for sample preparation (Qiagen).
13. Primers for Pyrosequencing (assay specific). *See Table 1* for primer sequences.
14. Plate for Pyrosequencing analysis, PyroMark Q96 Plate Low (Qiagen).

The wash, binding, and annealing buffers, denaturation solution, and streptavidin-coated Sepharose beads should be stored at 4 °C, the Pyrosequencing primers at –20 °C, and all other reagents at room temperature.

2.5 Analysis by Pyrosequencing Reaction

1. Pyrosequencing instrument: PyroMark Q96 ID System (Qiagen).
2. Cartridge for reagent dispensation: PyroMark Q96 Cartridge (Qiagen).
3. Pyrosequencing reagent kit: PyroMark Gold Q96 Reagents (Qiagen).
4. PyroMark Q96 ID Software 2.5 (Qiagen).
5. An additional alignment software, e.g., Gencious (Biomatters Ltd, Auckland, New Zealand), is optional, however yet recommended, as the PyroMark software has a very simplified version of alignment.

3 Methods

The protocol for molecular analysis of drug-resistant *M. tuberculosis* by Pyrosequencing can be divided into seven steps. The first two steps only need to be performed when setting up a new Pyrosequencing assay (*Setting up the assay*), and the subsequent

five steps are performed each time running the assay (*Performing the assay*). Below is a brief summary of the seven steps.

Setting up the assay

1. *Design of primers* for the PCR amplification and the sequencing reaction.
2. *Optimization* of the PCR and sequencing reactions.

Performing the assay

1. *The DNA extraction step*, to obtain DNA for the PCR reaction, preferably rendering pure and high-quality DNA.
2. *PCR amplification* of the DNA target performed with one of the PCR primers biotinylated.
3. *Sample preparation for the Pyrosequencing reaction*, which involves purification and denaturation of the PCR product into single-stranded DNA and subsequent annealing of the Pyrosequencing primer to the single-stranded DNA target.
4. *The automated Pyrosequencing reaction*, which synthesizes the complementary strand to the single-stranded DNA template which is analyzed and displayed in a Pyrogram®.
5. *Analysis* of the output sequence and the Pyrogram.

3.1 Assay Design

The *M. tuberculosis* genome has a mean G+C content of 65 % [21]. High GC content promotes secondary structure, such as hairpin loops. Secondary structure may inhibit binding of PCR and Pyrosequencing primers to the DNA template, which leads to inefficient PCR and sequencing reactions. A specific PCR product with good yield is essential for a successful result of Pyrosequencing. In other words, the assay design is a critical step in the whole procedure, and therefore, great effort should be paid to the assay design as it will influence the outcome of the assay.

There are several PCR assay design programs available, some free of charge. Users already accustomed with a specific program may be successful in using that for the design of PCR primers but should keep in mind that the sequencing reaction is run at a low temperature (28 °C) with single-stranded DNA. Pyrosequencing primers can be designed manually without the help of a software; however the commercially available PyroMark Assay Design is highly recommended as it gives valuable information on formation of duplexes and hairpin loops, and primer complementarity.

3.1.1 Design of PCR Primers

1. The length of the PCR primers should be between 18 and 24 nucleotides. Shorter primers may lead to nonspecific PCR amplification. Longer primers are more specific but have a higher probability of containing secondary structures (*see Note 4*).

2. The optimal length of the amplification product for Pyrosequencing assays is up to 250 bp, although products up to 500 bp may work well.
3. For PCR in general, primer GC content of 40–60 % is desirable; however, for sequences or genomes with high GC content (such as *M. tuberculosis*), primer GC content may be in the upper range or higher than 60 %. A melting temperature (T_m) preferably higher than 60 °C is recommended (*see Note 5*).
4. A T_m variation up to 5 °C is acceptable for different primers.
5. The choice, which primer is modified with a biotin label, will depend on the direction of the sequencing primer. The primer in the opposite direction of the sequence primer should be biotinylated for immobilization to the streptavidin-coated Sepharose beads (*see Note 6*).

3.1.2 Design of Pyrosequencing Primers

1. The design of the sequencing primer essentially follows the same rules as for the design of PCR primers except that the T_m of the sequencing primer may be lowered if necessary. This means that the sequencing primer can be made shorter than the PCR primers (*see Note 7*).
2. The distance between the sequencing primer and target region is optimally 0 and 3 nucleotides because the quality of the sequence diminishes with every nucleotide added to the reaction (*see Note 8*).
3. Homopolymers in the target sequence, especially nucleotide quadruples or more, should, if possible, be avoided as the peak height of the signal will be increasingly difficult to interpret in correlation to homopolymer length. In addition, this will likely reduce downstream sequencing quality.

3.1.3 Optimization of PCR and Pyrosequencing Reactions

1. A typical PCR program for PCR products up to about 250 bp, using a hot start DNA polymerase is as follows: initial denaturation and polymerase activation at 95 °C for 10 min; 45–50 cycles consisting of template denaturation 95 °C for 15 s, primer annealing at the annealing temperature for 20 s, elongation at 72 °C for 20 s; and a final elongation at 72 °C for 5–10 min (*see Note 9*).
2. An applicable annealing temperature is 5 °C below the calculated T_m of the primers. Optimize the annealing temperature by performing a gradient PCR or by testing a number of different temperatures over a range of 20 °C.
3. A typical primer concentration in the PCR reaction is 0.2 μ M (of each primer).
4. Perform a magnesium titration from 1 to 3 mM in 0.5 mM increments to determine the optimal magnesium concentration.

If the concentration is too low, no PCR product will be seen, and if too high concentration, an unspecific PCR product may be seen.

5. Check an aliquot of the PCR product on a 1.5 % agarose gel. There should be a clear, strong product band without excess of primers, primer-dimers, or other nonspecific products.
6. Perform Pyrosequencing reaction optimization by testing the following parameters individually: water, sequencing primers, and PCR primers, respectively. None of these reactions should give any sequence signal after successful optimization (*see Note 10*). Also test a good-quality PCR product with known sequence (reference sequence). Several different Pyrosequencing primers may need to be tested as well as different dispensation orders.

3.2 DNA Sample Preparation

We have been successful in performing Pyrosequencing on DNA extracted by the chloroform method described below; however, other standard procedures such as the phenol/chloroform method with or without proteinase K may also be successful. Preferably, the extracted DNA is pure and of high quality; however, a well-optimized assay has a higher tolerance for poor-quality DNA or for overloading of DNA.

1. Add 1–2 loops (10 μ L loops) of bacteria grown on solid medium to 750 μ L Tris–HCl in a screw-cap microcentrifuge tube.
2. Inactivate bacteria by placing the screw-cap tubes in an 85 °C water bath for 30 min.
3. Centrifuge for 5 min at 12,000 $\times g$.
4. Discard supernatant.
5. Add 50 μ L ultrapure water to the pellet.
6. Add 50 μ L chloroform to the pellet.
7. Vortex for 5 min.
8. Centrifuge for 5 min at 12,000 $\times g$.
9. Carefully remove the aqueous phase (top phase) and place it in a new microcentrifuge tube (screw cap not necessary). The aqueous phase contains water-soluble nucleic acids. Most cell debris will be present in the interphase between the aqueous and organic phases.
10. Measure the DNA concentration by a NanoDrop spectrophotometer and adjust the DNA concentration to 5–20 ng/ μ L.

3.3 PCR Amplification

A specific PCR product with good yield is preferred for a successful Pyrosequencing reaction. Too low yield will give a low signal, which could be difficult to discriminate from the background noise.

1. A quantity of 10–20 ng DNA is recommended for the PCR reaction.
2. Run the PCR reaction for 45–50 cycles. Fewer cycles may produce insufficient yield of template DNA and presence of unused biotinylated primer may interfere with subsequent Pyrosequencing reactions.

3.4 Sample Preparation for Pyrosequencing

In this step, the double-stranded PCR product from the previous step is rendered single stranded by immobilization of the biotinylated strand to the streptavidin-coated Sepharose beads and washing in sodium hydroxide (denaturation solution). Up to 96 samples can be prepared and sequenced at the same time. Several 96-well plates can be prepared sequentially. Pyrosequencing holds great flexibility; up to 96 different assays can be performed in parallel in one single run or one assay can analyze 96 samples simultaneously.

1. Prepare a master mix containing 40 μL binding buffer, 3 μL Sepharose beads, and 7 μL water for each sample.
2. Distribute 50 μL of the master mix into a new PCR plate, and then add 30 μL of the PCR product to each well.
3. Cover the plate with sealing tape and incubate the mixture for 10 min at room temperature under constant mixing (Make sure that the solution is thoroughly mixed without any solution touching the sealing tape) (*see Note 11*).
4. During mixing in **step 3**, fill the prep station troughs with water, ethanol, denaturation solution, and wash buffer. Fill troughs with approximately 180 mL of the respective solution, except for the trough for denaturation solution, which should be filled with approximately 120 mL (*see Note 12*). Rinse the filter tips with water. Set the heating device on 80 °C.
5. During mixing in **step 3**, dilute the sequencing primer in annealing buffer to a concentration of 0.4 μM and transfer 40 μL to a sequencing plate (PyroMark Q96 Plate Low). Place the sequencing plate into its position on the vacuum prep worktable. Several different primers can be used in the same plate; that is, several assays can be performed in the same sequencing plate.
6. Within 1 min from stopping agitation, capture the beads (with immobilized PCR products) to the filter probes by lowering the vacuum prep tool (vacuum turned on) into the PCR plate. Avoid moving filter tips horizontally to risk losing Sepharose beads by contacting the walls of the PCR plate.
7. Keeping the vacuum turned on, move the prep tool into the ethanol trough, let wash for 5 s, and then move to denaturation solution for 5 s, and finally to the washing buffer for 10 s (*see Note 13*).
8. The vacuum must be turned off before placing the vacuum prep tool in the sequencing plate, as otherwise the annealing

buffer with the sequencing primer will be flushed through the prep tool. Turn the vacuum off and release the beads by gently shaking the vacuum prep tool from side to side in the sequencing plate. The lower part of each filter probe should be fully merged into the pre-loaded annealing buffer with primers.

9. Move the prep tool to the water trough and wash the filter probes for 10 s (vacuum turned on).
10. Heat the sequencing plate at 80 °C for 2 min (*see* **Note 14**).
11. Remove the plate from the heating block and let samples cool to room temperature.

3.5 The Pyrosequencing Reaction

1. Prepare the run in PyroMark Q96 ID Software 2.5.
2. Resuspend the enzyme and substrate mixtures with ultrapure water and shake the glass bottle gently a couple of times to stir (do not vortex). Incubate at room temperature for 5–10 min.
3. Rinse the cartridge and spray the outside of the needles using ultrapure water. Fill the compartments to the top with ultrapure water and check for obstruction by pressing on top of each compartment with a finger (using clean gloves). Ensure that all six needles are clear and not deformed; that is, a straight downward directed beam of water should be visible from the tip of the needle when checking for obstruction. Empty the cartridge by turning it upside down and gently shaking it.
4. Pipette enzyme and substrate mixtures and dNTPs (volumes are given when finalizing the protocol in the PyroMark Q96 ID software) into their designated positions in the cartridge.
5. Place the sequencing plate and the filled cartridge in the PyroMark Q96 ID System.
6. Start the run.
7. After the run, rinse each compartment of the cartridge four times as well as spray the outside of the needles using ultrapure water. Fill the compartments to the top with ultrapure water and check for obstruction by pressing on top of the compartment with a finger as in **step 3**. Ensure that the needles are clear; that is, a beam of water should be visible at the tip of the needle. Empty the cartridge by turning it upside down. Place the cartridge on the side, on a lint-free tissue, and leave to dry.

3.6 The Analysis

1. In the PyroMark Q96 ID Software 2.5, click “Analyze all wells” or “Analyze the selected wells only.”
2. Export Pyrograms and FASTA files from the PyroMark Q96 ID Software 2.5.
3. Perform analysis of the sequences by aligning with a reference sequence. Verify sequences by inspecting the Pyrograms (*see* **Note 15**).

4 Notes

1. A successful Pyrosequencing assay is dependent on the efficiency of the PCR and quality of the sequencing reaction. Mycobacterial DNA has a high content of Gs and Cs compared to many other bacteria. PCR amplification and subsequent sequencing of GC-rich DNA are often problematic due to stable secondary structures in the DNA that are resistant to melting and may interfere with primer hybridization. Primer-dimer and hairpin loops may also be problematic. Therefore, effort should be put into designing PCR and sequencing primers.
2. Biotinylated primers should always be ordered HPLC purified to assure correct primer size and removal of free biotin which competes with the PCR product for the binding sites of the streptavidin-coated beads used for the template preparation. It should also be noted that the quality of biotinylated primers may differ from manufacturer to manufacturer. Biotinylated primers are particularly sensitive to storage; therefore it is important to store them at -20°C , not at 4°C .
3. The different buffers used for template preparation can also be prepared in-house. Binding buffer contains 10 mM Tris-HCl, 2 M NaCl, 1 mM EDTA, and 0.1 % Tween 20, pH 7.6. Denaturing solution: 0.2 M NaOH. Washing buffer: 10 mM Tris-acetate, pH 7.6. Annealing buffer: 20 mM Tris-acetate, and 2 mM Mg-acetate, pH 7.6.
4. The primers should not form heavy hairpin loops or dimers with themselves or the other primer. Check the biotinylated PCR primer with great care for hairpin loops and duplexes, as excess of biotinylated primer might cause background in the Pyrosequencing assay.
5. The strand separation temperature for GC-rich DNA is notably higher than for DNA with lower GC content. At lower temperatures, the two strands of PCR amplification products have a tendency to reanneal faster and, as a result, compete with primer annealing. Higher annealing temperatures during PCR favor primer annealing and therefore increase amplification efficiency.
6. If the forward primer is biotinylated, the sequence of the minus strand will be obtained. If the reverse primer is biotinylated, the sequence of the plus strand will be obtained.
7. It is crucial to check for primer-dimers and hairpins since the Pyrosequencing reaction is run at 28°C .
8. Directed nucleotide dispensation order instead of cyclic dispensation order may increase sequence read length.

9. For best yield, and consumption of all biotinylated PCR primers, which is important for Pyrosequencing, 45–50 PCR cycles should be run when using a hot start DNA polymerase compared to 35–50 cycles when using ordinary DNA polymerase.
10. Self-complementary hairpin loops of the 3' end of the biotinylated single-stranded DNA template are a recognized problem in Pyrosequencing assay design. If present, this will be detected as a recognizable sequence when running a full sequencing reaction without any sequencing primers. Many times this self-priming can be circumvented by adding non-sense nucleotides to the 5' end of the non-biotinylated primer.
11. The Sepharose beads sediment quickly; therefore the sample plate must be kept under constant mixing, or at least 10 min before starting sample preparation. Sample preparation should be initiated within 1 min of stopping agitation.
12. Note that the volume of denaturation solution must be less than the wash buffer in order to avoid transfer of denaturation solution (on the filter probes) to the Pyrosequencing reaction samples. Denaturation solution may interfere with the Pyrosequencing primer binding to its target.
13. Move the tool beyond 90° vertical after picking it up to allow liquid to completely drain from the filter probes. Move back to a horizontal position and place in the next trough.
14. For some very GC-rich templates, it may help to increase the heating to approximately 90 °C for 2 min, or to prolong the time at 80 °C. In this case, make sure to seal the plate to reduce evaporation when using annealing protocols with extended heating time and/or elevated temperature.
15. Note that the PyroMark Q96 ID software may misinterpret peaks in the Pyrogram, although fairly unusual. The algorithm generally misinterprets the same peak in different samples. It may therefore be a good practice to verify mutations or unexpected sequences manually in the Pyrogram.

References

1. World Health Organization (2010) Multidrug and extensively drug-resistant TB (M/XDR-TB): 2010 global report on surveillance and response. vol WHO/HTM/TB/2010.3. Geneva, Switzerland
2. Bergval I, Sengstake S, Brankova N et al (2012) Combined species identification, genotyping, and drug resistance detection of *Mycobacterium tuberculosis* cultures by MLPA on a bead-based array. PLoS One 7: e43240
3. Bergval IL, Vijzelaar RN, Dalla Costa ER et al (2008) Development of multiplex assay for rapid characterization of *Mycobacterium tuberculosis*. J Clin Microbiol 46:689–699
4. Hernandez-Neuta I, Varela A, Martin A et al (2010) Rifampin-isoniazid oligonucleotide typing: an alternative format for rapid detection of multidrug-resistant *Mycobacterium tuberculosis*. J Clin Microbiol 48:4386–4391
5. Hillemann D, Rusch-Gerdes S, Richter E (2007) Evaluation of the GenoType MTBDRplus assay for rifampin and isoniazid susceptibility testing of *Mycobacterium tuberculosis* strains and clinical specimens. J Clin Microbiol 45:2635–2640

6. Hillemann D, Rusch-Gerdes S, Richter E (2009) Feasibility of the GenoType MTBDRsl assay for fluoroquinolone, amikacin-capreomycin, and ethambutol resistance testing of *Mycobacterium tuberculosis* strains and clinical specimens. *J Clin Microbiol* 47:1767–1772
7. Helb D, Jones M, Story E et al (2010) Rapid detection of *Mycobacterium tuberculosis* and rifampin resistance by use of on-demand, near-patient technology. *J Clin Microbiol* 48:229–237
8. Guo Q, Zheng RJ, Zhu CT et al (2013) Pyrosequencing for the rapid detection of rifampicin resistance in *Mycobacterium tuberculosis*: a meta-analysis. *Int J Tuberc Lung Dis* 17:1008–1013
9. Campbell PJ, Morlock GP, Sikes RD et al (2011) Molecular detection of mutations associated with first- and second-line drug resistance compared with conventional drug susceptibility testing of *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* 55:2032–2041
10. Engström A, Morcillo N, Imperiale B et al (2012) Detection of first- and second-line drug resistance in *Mycobacterium tuberculosis* clinical isolates by pyrosequencing. *J Clin Microbiol* 50:2026–2033
11. Arnold C, Westland L, Mowat G et al (2005) Single-nucleotide polymorphism-based differentiation and drug resistance detection in *Mycobacterium tuberculosis* from isolates or directly from sputum. *Clin Microbiol Infect* 11:122–130
12. Bravo LT, Tuohy MJ, Ang C et al (2009) Pyrosequencing for rapid detection of *Mycobacterium tuberculosis* resistance to rifampin, isoniazid, and fluoroquinolones. *J Clin Microbiol* 47:3985–3990
13. Jureen P, Engstrand L, Eriksson S et al (2006) Rapid detection of rifampin resistance in *Mycobacterium tuberculosis* by Pyrosequencing technology. *J Clin Microbiol* 44:1925–1929
14. Kalokhe AS, Shafiq M, Lee JC et al (2013) Multidrug-resistant tuberculosis drug susceptibility and molecular diagnostic testing. *Am J Med Sci* 345:143–148
15. Ronaghi M, Karamohamed S, Pettersson B et al (1996) Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem* 242:84–89
16. Scorpio A, Zhang Y (1996) Mutations in *pncA*, a gene encoding pyrazinamidase/nicotinamidase, cause resistance to the antituberculous drug pyrazinamide in *tubercle bacillus*. *Nat Med* 2:662–667
17. Maus CE, Plikaytis BB, Shinnick TM (2005) Mutation of *tlyA* confers capreomycin resistance in *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* 49:571–577
18. Engström A, Perskvist N, Werngren J et al (2011) Comparison of clinical isolates and in vitro selected mutants reveals that *tlyA* is not a sensitive genetic marker for capreomycin resistance in *Mycobacterium tuberculosis*. *J Antimicrob Chemother* 66:1247–1254
19. Lin SY, Rodwell TC, Victor TC et al (2014) Pyrosequencing for rapid detection of extensively drug-resistant *Mycobacterium tuberculosis* in clinical isolates and clinical specimens. *J Clin Microbiol* 52:475–482
20. Dunker J, Larsson U, Petersson D et al (2003) Parallel DNA template preparation using a vacuum filtration sample transfer device. *Biotechniques* 34(862–866):868
21. Camus JC, Pryor MJ, Medigue C et al (2002) Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology* 148:2967–2973

Chapter 25

Application of Pyrosequencing® in Food Biodefense

Kingsley Kwaku Amoako

Abstract

The perpetration of a bioterrorism attack poses a significant risk for public health with potential socioeconomic consequences. It is imperative that we possess reliable assays for the rapid and accurate identification of biothreat agents to make rapid risk-informed decisions on emergency response. The development of advanced methodologies for the detection of biothreat agents has been evolving rapidly since the release of the anthrax spores in the mail in 2001, and recent advances in detection and identification techniques could prove to be an essential component in the defense against biological attacks. Sequence-based approaches such as Pyrosequencing®, which has the capability to determine short DNA stretches in real time using biotinylated PCR amplicons, have potential biodefense applications. Using markers from the virulence plasmids and chromosomal regions, my laboratory has demonstrated the power of this technology in the rapid, specific, and sensitive detection of *B. anthracis* spores and *Yersinia pestis* in food. These are the first applications for the detection of the two organisms in food. Furthermore, my lab has developed a rapid assay to characterize the antimicrobial resistance (AMR) gene profiles for *Y. pestis* using Pyrosequencing. Pyrosequencing is completed in about 60 min (following PCR amplification) and yields accurate and reliable results with an added layer of confidence, thus enabling rapid risk-informed decisions to be made. A typical run yields 40–84 bp reads with 94–100 % identity to the expected sequence. It also provides a rapid method for determining the AMR profile as compared to the conventional plate method which takes several days. The method described is proposed as a novel detection system for potential application in food biodefense.

Key words *Bacillus anthracis*, *Yersinia pestis*, Antimicrobial resistance, Pyrosequencing®

1 Introduction

Food is reported to be a vulnerable target for intentional contamination and the use of *Salmonella enterica* serovar Typhimurium in the tainting of salad bars in the USA makes the use of biothreat agents such as *B. anthracis* spores a possibility [1]. The prospect of this vulnerability is even more alarming given that no standards or established guidelines exist for testing foods for these biothreat agents [2]. Unambiguous and accurate information obtained within a relevant time frame is required to make a rapid risk-informed decision [3]. The ability to generate DNA sequence data

from a suspected biothreat agent provides an added layer of confidence compared to a presumptive positive PCR amplicon on a gel; hence sequencing-based technologies, such as Pyrosequencing[®], have sufficient discrimination potential to be used for microbial identification and can also be used to identify the presence of anti-microbial resistance (AMR) genes [3].

Among the genotypic techniques used for microbial detection and typing, only the sequencing-based techniques possess the highest resolution and accuracy required. Pyrosequencing fulfills the aforementioned criteria for use in biodefense for microbial detection, identification, and typing. Pyrosequencing is a sequencing-by-synthesis method that quantitatively monitors the incorporation of nucleotides in real time, through the emission of light following the enzymatic conversion of pyrophosphate released during nucleotide incorporation [4]. It generates similar data to Sanger sequencing [5] and is a rapid, reproducible, high-throughput, user-friendly, and cost-effective method [6, 7]. Pyrosequencing produces nucleotide sequence information and is more reliable compared to PCR alone and the data handling is simple and reproducible for comparison between laboratories and potential epidemiological characterization of isolates [3]. Even though Pyrosequencing is limited to short fragments, the careful selection of targets can provide enough information for accurate identification and typing of biothreat agents [3]. Several reports have shown the use of Pyrosequencing in a wide range of applications including SNR analysis (Ref. 8, *see also* Chapter 26), SNP typing [9, 10], identification and subtyping based on the 16S rRNA gene [11], identification of antibiotic resistance genes [12–16] and mutations conferring resistance to antibiotics (Ref. 17, *see also* Chapter 24), diagnosis of human disease (Ref. 18, *see also* Chapters 5–8, 10), determination of gene copy number (Ref. 19, *see also* Chapter 9), development of a molecular Gram stain [20], and metagenomics [21].

Pyrosequencing has potential merit in biodefense in providing a rapid sequence-based identification tool and hence an added layer of confidence in the analysis of biothreat agents. The *Bacillus cereus* group including *B. anthracis* [22] shares a great deal of morphological, biochemical, and genetic similarities [23–25], making differentiation an arduous task. *Bacillus anthracis*, implicated in the 2001 anthrax attack in the USA [26], has received wide attention as a bioterrorism agent. Several reports have explored the use of different methods for the detection and genotypic characterization of *B. anthracis* [22, 25, 27–34]; however most of these are limited in discriminating between genetic similarities [35]. Considering that these approaches for species identification can be tedious, expensive, and inaccurate, a rapid and unambiguous method yielding sequence information is needed. With the increasing use of sequencing-based methods and decreased expense of running

sequencing reactions after the initial equipment investment, more laboratories are relying on sequence data for species identification [36]. Pyrosequencing technology which has the capability to determine short DNA stretches in real time using biotinylated PCR amplicons [4] has recently been used for rapid sequence-based detection in several potential biodefense applications [3, 37–40]. The application of Pyrosequencing for the detection of *B. anthracis* has been previously reported [40, 41]. Recently, my lab has reported the first application of Pyrosequencing for the specific detection and confirmation of *B. anthracis* and *Y. pestis* from food [37, 38]. More recently, we have also designed TaqMan probes for the Pyrosequencing targets and demonstrated the combined use of real-time PCR with Pyrosequencing for rapid detection and quantitation [42]. With this new assay, we run real-time PCR on the extracted DNA instead of just PCR and use the amplicons generated from the real-time PCR for Pyrosequencing. The Pyrosequencing assays designed are based on chromosomal and virulence markers and in combination with an immunomagnetic assay previously developed [43], the work demonstrates excellent detection capability with greater accuracy as evidenced by the high sequence identities (Table 1 and Fig. 1). Although AMR in *Y. pestis* is rare, the emergence and spread of multidrug-resistant strains, under natural or engineered conditions, could have catastrophic

Table 1

Pyrosequencing read lengths and BLAST identities for *B. anthracis* Sterne spores in the food matrices

Food matrix	Target	Read length (bp)	% Identity (BLAST)
Water	cya	77	100
	gerXB	80	>94
	Prophage Lambda 1	67	>95
	Prophage Lambda 3	72	>97
Apple juice	cya	77	100
	gerXB	80	>96
	Prophage Lambda 1	67	>95
	Prophage Lambda 3	72	>97
Milk	cya	77	100
	gerXB	80	>96
	Prophage Lambda 1	67	>97
	Prophage Lambda 3	72	>97
Sliced ham	cya	77	100
	gerXB	80	>96
	Prophage Lambda 1	67	>97
	Prophage Lambda 3	72	>98

From Amoako et al. 2013 [38]

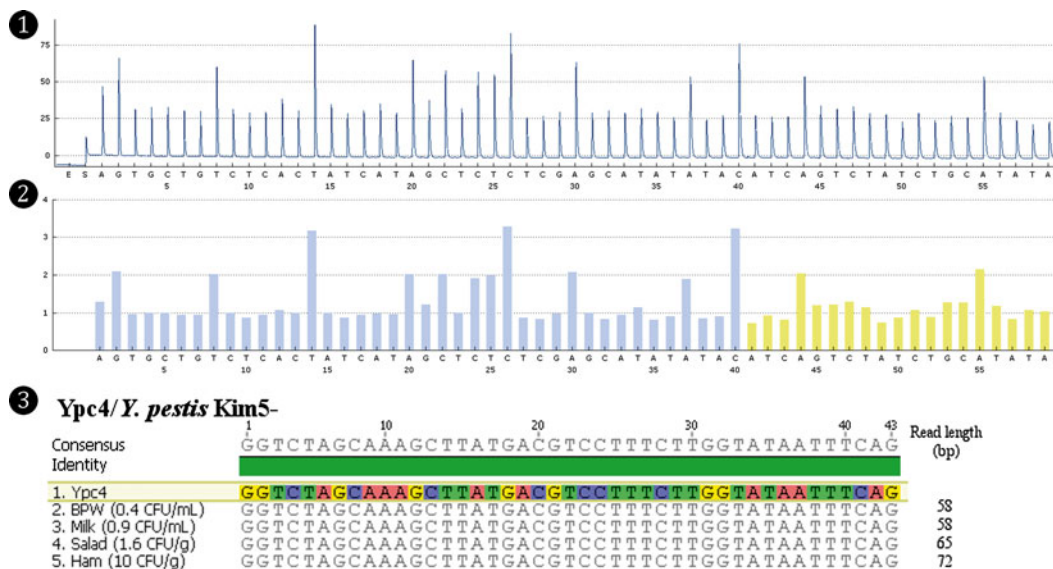


Fig. 1 Pyrosequencing data showing the raw Pyrogram® (1), base calls based on peak height along with quality scores (2), and an alignment view of the sequence data in relation to the reference strain *Y. pestis* C092 (3)

consequences [3]. My lab has recently demonstrated the application of Pyrosequencing in the rapid identification of virulent genes and AMR genes in *Y. pestis* [3]. The technology has potential application in biodefense and hence protocols and standard operating procedures developed for this specific application in potential food biodefense response are described in this chapter.

2 Materials

2.1 Personal

Protective Equipment

1. BSL-2.
2. Standard lab coat and latex or nitrile gloves, long pants, close-toed shoes.

2.2 Design

of Primers for

PyroMark® Assays

1. Geneious Software Suite (version 5.3.5; Biomatters Inc).
2. PyroMark Assay Design software.
3. NCBI Primer-BLAST tool (<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>).

2.3 Extraction

of Genomic DNA

MasterPure Gram Positive DNA Purification Kit (Epicenter Biotechnologies, Madison, WI, USA).

2.4 PCR Amplification

PCR amplification was performed as previously described [[3](#), [37](#), [38](#)].

1. PyroMark PCR master mix 2× (Qiagen).
2. RNase-free water.
3. Oligonucleotide primers (*see* Table [2](#)).
4. Template DNA.
5. 200 µL reaction tubes (strips or single-capped tubes).
6. Pipettor and aerosol-resistant pipet tips.
7. Thermal cycler.

2.5 Pyrosequencing Analysis

1. PyroMark Q24 software.
2. PyroMark Gold Q24 reagents, containing:
 - (a) Enzyme mixture.
 - (b) Substrate mixture.
 - (c) dATPαS.
 - (d) dGTP.
 - (e) dCTP.
 - (f) dTTP.
3. 70 % reagent-grade ethanol.
4. PyroMark Binding buffer (Qiagen).
5. Streptavidin coated Sepharose beads (Thermo Fisher Scientific).
6. Sequencing primer (*see* Table [2](#)).
7. PyroMark Annealing buffer (Qiagen).
8. PyroMark Denaturation buffer (Qiagen).
9. PyroMark Wash buffer (Qiagen).
10. Milli-Q water.
11. Pipettor and aerosol-resistant pipet tips.
12. Snap apart 96-well plates.
13. PyroMark Q24 plate (Qiagen).
14. PyroMark vacuum preparation workstation.
15. PyroMark vacuum pump.
16. PyroMark workstation troughs (Qiagen).
17. Heat block.
18. PyroMark Q24 plate holder (Qiagen).
19. Q24 Cartridge (Qiagen).

2.6 Data Analysis

1. PyroMark Q24 Software.

Table 2
PCR primers and Pyrosequencing primers for the detection of *Bacillus anthracis* and *Yersinia pestis*.^{a, b}

Target	Gene	Primer sequence (5'-3')	Amplicon size (bp)	Sequencing primer (5'-3')
Ba pXO1	<i>cya</i>	F:GTGGTGTGGCTACAAAGGGATTGAATGT	123	CAAAAGGGATTGAATGTT
		Biotin-R:TCTCGACAGCTAATTGTTGACCATGCTTCT		
	<i>gerXB</i>	F:TGGAGTTTGGTGCCATTGTGGCCG	129	GTGGCCGCAAAATCAG
		Biotin-R:AAACCCCTACAAGCCCACTGGTACAC		
Ba pXO2	<i>acpB</i>	Biotin-F:CACAGGAGAAAGTGAAGTTGGGCAGA	166	AGTCAGCATTAACATCGT
	<i>capBCAD</i>	R:AGGGGTGAAACCTAAGTCAAGAGGTATTGT		
		Biotin-F:GTCTTGCAAGCTACCCCTTACCATGG	139	Same as reverse primer
		R:ACGAACGTTTATGGCCCCCACTGTAT		
Ba Chromosome	<i>Prophage lambda 1</i>	F:GGGGGATTAAATTGCAAAAGCGTGTT	94	TTGCAAAAAGCGTGTT
	<i>Prophage lambda 3</i>	Biotin-R:CCCATGGCGTCCCATAGTATGA		
		F:GGTGATGCGTCAAAAGCCGATGGA	122	AAAAGCCGATGGAAC
		Biotin-R:TTACGTGGGCCCAAATGTCCCGTTT		
Yp Chromosome	<i>Hypothetical protein CDS</i>	F:CGCAAAACTCCTCAACTCTACATA	152	TGCTCGTGATTATGTGT
		Biotin-R:TGAGGATTATATTCCCGCTATT		
Yp pMT1	<i>caf1M</i>	GGAATTATTACTTTTCGGCATGCTT	135	TCGGCATGCTTAGTTTTT
		Biotin-R:CATAACGCCAGCAGCATCTA		
Yp pPCP1	<i>pst</i>	F:AGCTTCATTATCAGAAAGCGTCAG	146	TCTCACTTTATCATAAGCCT
		Biotin-R:TTGGTCAGCGAAGCAAAACA		

^aPrimer sets previously published as:

Amoako KK, Janzen TW, Shields MJ et al. (2013) Rapid detection and identification of *Bacillus anthracis* in food using pyrosequencing technology. *Int J Food Microbiol* 165:319–325

Amoako KK, Thomas MC, Kong F, et al. (2012) Rapid detection and antimicrobial resistance gene profiling of *Yersinia pestis* using pyrosequencing technology. *J Microbiol Methods* 90:228–234

^bThermal cycling conditions include enzyme activation at 95 °C for 5 min followed by 50 cycles of denaturing at 95 °C for 30 s, annealing at 58 °C (60 °C for Yp targets) for 30 s, and extension at 72 °C for 30 s

3 Methods

3.1 Design of Primers for PyroMark Assays

Consensus sequences were determined using the Geneious Software Suite as previously described [3, 37, 38]. The consensus sequence targets were imported into the PyroMark Assay Design software, and used to design the sequencing primers based on the sequence analysis method (SQA) for all targets (Table 2). All primers generated were examined for specificity *in silico* using the NCBI Primer-BLAST tool (<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>).

1. Launch the PyroMark Assay Design software.
2. Select File→New→Sequence Analysis (SQA).
3. Paste the gene sequence of interest into the Sequence Editor.
4. Highlight the sequence of interest, then right-click, and select Target Region→Select Target Region.
5. Press F5 on your keyboard to start the assay design process.
6. Right-click on a primer set on the right side of the window and select “View Report” (*see Note 1*).
7. Save the report file including designed primers (as html, page only) by clicking the save button (i.e., diskette image).
8. Close the saved report and save the assay design by selecting File→Save As.

3.2 Extraction of Genomic DNA

Genomic DNA from *B. anthracis* was extracted as previously reported according to the manufacturer’s instructions, using the MasterPure Gram Positive DNA Purification Kit [38]. Genomic DNA was obtained from *Y. pestis* extracted from food samples by using a simple boiling lysis method described in detail elsewhere [44].

3.3 PCR Amplification

1. Thaw the PyroMark PCR master mix and primers at room temperature.
2. Combine PCR components according to Table 3.
3. Gently mix the master mix by vortexing and dispense the appropriate volume into PCR tubes.
4. Add template DNA (~10 ng) to PCR tubes (*see Note 2*).
5. Program the thermal cycler according to instructions in Table 2.
6. Place the PCR tubes in the thermal cycler and start program.
7. Run samples using QIAxcel protocol.
8. Use 5–20 µL of a 25 µL PCR for subsequent Pyrosequencing analysis (*see Note 3*).

3.4 Pyro Run Assay Setup

1. Launch the PyroMark Q24 software.
2. Select File→New Assay→SQA Assay.
3. Enter the desired dispensation order.
4. Select File→Save As and save the dispensation.
5. From the file tab select new run (SQA).
6. Select instrument method from the drop-down tab (*see Note 4*).
7. Under the plate setup, right-click a well and select load assay. Select appropriate assay created in **step 4**. Repeat as necessary for each assay you would like to add to the plate.
8. In the bottom two sections of each well, enter well and sample information.
9. Under the tools tab, select show protocol setup. Print this sheet off as it contains all the information needed to fill the reagent cartridge for the Pyrosequencing run.
10. Save the run file onto a USB drive and proceed to Subheading 2.5.

3.5 Pyrosequencing Run

3.5.1 Immobilizing the PCR Products

1. Make up master mix according to Table 4.
2. Dispense 60–75 μL of the master mix (Table 3) into up to 24 wells of a snap apart 96-well plate and add 5–20 μL of each PCR product to the appropriate well.
3. Place the plate on an orbital shaker at $4,403\times g$ for 5–10 min.
4. While the amplicon is binding to the beads prepare the sequencing primer master mix as described in Table 5.
5. Aliquot 25 μL to each well of the PyroMark Q24 plate.
6. Fill the workstation troughs as shown in Fig. 2.
7. Flush the filter probes by starting the vacuum pump, opening the valve on/off switch, and placing unit in trough 5 containing high-purity water. Refill with 70 mL of high-purity water for later use.

Table 3
Qiagen PyroMark PCR setup

Reagent	Volume/reaction	Final concentration
PyroMark PCR master mix (2 $\times$)	12.5 μL	1 $\times$
Primer F	Variable	0.2 μM
Primer R	Variable	0.2 μM
RNase-free water	Variable	Variable
Template DNA (5 ng/ μL)	2 μL	Added to each tube
Total volume 25 μL		

Table 4**Master mix for binding biotinylated PCR products to streptavidin-coated beads**

Reagent	Volume/reaction (μL)	Final concentration
Beads	2	NA
Binding buffer	40	1×
RNase-free water	18–33	NA
PCR product	5–20	Added separately to each tube
Total volume = 80 μL		

Table 5**Master mix for annealing of the sequencing primer**

Reagent	Volume/reaction	Final concentration
Annealing buffer	24.25	1×
Sequencing primer (10 μM)	0.75	0.3 μM
Total volume = 25 μL		



Fig. 2 PyroMark Q24 setup including PyroMark instrument (*right*) and vacuum workstation (*left*) showing troughs containing 50 mL of 70 % ethanol (1), 40 mL denaturation solution (2), 50 mL wash buffer (3), and high-purity water (4 and 5 containing 50 mL and 70 mL, respectively)

8. Place the PCR plate from **step 3** on the workstation in the mold below trough 1. Retain the proper orientation in which the bead/PCR product mixture was added.
9. Turn the vacuum switch on and lower the vacuum tool into the PCR plate wells. Hold in position for 15 s to ensure complete capture of streptavidin-coated beads (*see Note 5*).
10. Move the tool to trough 1 for 5 s.
11. Move the tool to trough 2 for 5 s.
12. Move the tool to trough 3 for 10 s.
13. Maintain the vacuum and raise the vacuum tool such that the pins are parallel to the ground for 5 s and then turn the vacuum OFF.
14. Align the vacuum tool with the wells in the PyroMark Q24 plate and lower the tool into the wells. Gently shake from side to side to release the beads.
15. Move the vacuum tool to trough 4 and shake for 10 s.
16. Move the vacuum tool to trough 5, switch the vacuum ON, and flush the filter probes with high-purity water for 5 s.
17. Raise the vacuum tool such that it is perpendicular to the ground for 5 s, then switch the vacuum OFF, and store in the parking position (above trough 5).
18. If done with all buffers and reagents for the day, discard the remaining liquid in the troughs using appropriate procedures (*see Note 6*).

3.5.2 Annealing of Sequencing Primers to DNA Strand

1. Move the PyroMark Q24 plate onto a heat block containing the PyroMark Q24 plate holder at 80 °C for 2 min.
2. Remove PyroMark Q24 plate and incubate at room temperature for at least 5 min.
3. Fill the PyroMark reagent cartridge with enzyme, substrate, and nucleotides as per the instructions on the sheet printed in Subheading 2.4 (step 9).
4. Place the plate on the PyroMark Q24 along with the reagent cartridge and insert the USB stick containing the run file.
5. Locate the run file on the PyroMark Q24 screen and click run.

3.6 Data Analysis

1. Open the run file using the PyroMark Q24 software.
2. Select the Analysis Mode (SQA in this example).
3. Select Analyze all.
4. Select individual wells to manually view the sequence data.
5. Export the data as a .fasta file by selecting Tools→Export as FASTA (*see Note 7*).
6. Blast and/or align the sequences to the reference sequence as shown in Fig. 1.

4 Conclusion

In conclusion, the method described demonstrates the power of Pyrosequencing in the detection of *B. anthracis* and *Y. pestis* from food and the application for AMR gene profiling in *Y. pestis*. The method provides a novel detection system with a potential application in a foodborne bioterrorism response. The Pyrosequencing method described is a rapid and reproducible technique yielding high sequence identity for the unambiguous identification of *B. anthracis* and *Y. pestis*. The 1-h time for completing Pyrosequencing runs following PCR amplification offers an effective application for food biodefense. Furthermore, a novel Pyrosequencing method for AMR gene profiling that provides an alternative method for determining potential antimicrobial resistances in *Y. pestis* has been described. The methods offer enhanced confidence in the positive identification of *B. anthracis* and *Y. pestis* over conventional DNA-based methods through the generation of sequence data.

5 Notes

1. High-scoring primers typically perform very well; however, low-scoring primers may also work well. Primers with low scores are worth investigating if no others are available.
2. These tubes contain concentrated template DNA. Extreme caution must be used when handling these tubes and their contents so as not to contaminate equipment, reagents, or other samples. This includes avoiding the creation of aerosols. Under no circumstances are these tubes or their contents to be used/handled in the PCR reagent preparation area.
3. This kit is also supplied with Coral load, which is not required in the reaction. Q solution and MgCl₂ are also provided for PCR with difficult templates. Avoid use of Q solution and MgCl₂ unless it is explicitly stated to use in Table 2. If optimizing a new primer set and no product is observed or the product is weak refer to Qiagen PyroMark PCR manual for instruction on use of Q solution and/or MgCl₂.
4. The instrument method revision is listed on the reagent cartridge (e.g., rev 003).
5. Keep the vacuum tool on from **steps 10 to 14**.
6. Refer for details in the PyroMark Q24 user manual.
7. Use the default parameters: Wells = all; Sort by = row; Included sequence = all.

Acknowledgements

Thank you to all the personnel in the Amoako lab who provided technical assistance in the work on the application of Pyrosequencing for the detection of *Bacillus anthracis* and *Yersinia pestis* in food. The assistance of Matthew Thomas in compiling this work is acknowledged. This work was funded by the Defence Research Development Canada Centre for Security Science, Chemical, Biological, Radiological, Nuclear and Explosive Research Technology Initiative (CRTI) funding from CRTI 08-0203RD.

References

1. Torok TJ, Tauxe RV, Wise RP et al (1997) A large community outbreak of salmonellosis caused by intentional contamination of restaurant salad bars. *JAMA* 278:389–395
2. Kennedy S (2008) Epidemiology. Why can't we test our way to absolute food safety? *Science* 322:1641–1643
3. Amoako KK, Thomas MC, Kong F et al (2012) Rapid detection and antimicrobial resistance gene profiling of *Yersinia pestis* using pyrosequencing technology. *J Microbiol Methods* 90:228–234
4. Ronaghi M, Karamohamed S, Pettersson B et al (1996) Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem* 242:84–89
5. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74:5463–5467
6. Ronaghi M (2001) Pyrosequencing sheds light on DNA sequencing. *Genome Res* 11:3–11
7. Ronaghi M, Elahi E (2002) Pyrosequencing for microbial typing. *J Chromatogr B Analyt Technol Biomed Life Sci* 782:67–72
8. Hahn KR, Janzen TW, Thomas MC et al (2014) Single nucleotide repeat analysis of *B. anthracis* isolates in Canada through comparison of pyrosequencing and Sanger sequencing. *Vet Microbiol* 169:228–232
9. Alderborn A, Kristofferson A, Hammerling U (2000) Determination of single-nucleotide polymorphisms by real-time pyrophosphate DNA sequencing. *Genome Res* 10:1249–1258
10. Unnerstad H, Ericsson H, Alderborn A et al (2001) Pyrosequencing as a method for grouping of *Listeria monocytogenes* strains on the basis of single-nucleotide polymorphisms in the *inlB* gene. *Appl Environ Microbiol* 67:5339–5342
11. Jonasson J, Olofsson M, Monstein HJ (2002) Classification, identification and subtyping of bacteria based on pyrosequencing and signature matching of 16S rDNA fragments. *APMIS* 110:263–272
12. Bravo LT, Tuohy MJ, Ang C et al (2009) Pyrosequencing for rapid detection of *Mycobacterium tuberculosis* resistance to rifampin, isoniazid, and fluoroquinolones. *J Clin Microbiol* 47:3985–3990
13. Arnold C, Westland L, Mowat G et al (2005) Single-nucleotide polymorphism-based differentiation and drug resistance detection in *Mycobacterium tuberculosis* from isolates or directly from sputum. *Clin Microbiol Infect* 11:122–130
14. Gharizadeh B, Akhras M, Unemo M et al (2005) Detection of *gyrA* mutations associated with ciprofloxacin resistance in *Neisseria gonorrhoeae* by rapid and reliable pre-programmed short DNA sequencing. *Int J Antimicrob Agents* 26:486–490
15. Naas T, Poirel L, Nordmann P (2006) Pyrosequencing for rapid identification of carbapenem-hydrolysing OXA-type β -lactamases in *Acinetobacter baumannii*. *Clin Microbiol Infect* 12:1236–1240
16. Borman AM, Petch R, Linton CJ et al (2008) *Candida nivariensis*, an emerging pathogenic fungus with multidrug resistance to antifungal agents. *J Clin Microbiol* 46:933–938
17. Haanperä M, Huovinen P, Jalava J (2005) Detection and quantification of macrolide resistance mutations at positions 2058 and 2059 of the 23S rRNA gene by pyrosequencing. *Antimicrob Agents Chemother* 49:457–460
18. Kruckeberg KE, Thibodeau SN (2004) Pyrosequencing technology as a method for the diagnosis of multiple endocrine neoplasia type 2. *Clin Chem* 50:522–529
19. Söderbäck E, Zackrisson AL, Lindblom B et al (2005) Determination of CYP2D6 gene copy number by pyrosequencing. *Clin Chem* 51:522–531

20. Kobayashi N, Bauer TW, Tuohy MJ et al (2006) The comparison of pyrosequencing molecular Gram stain, culture, and conventional Gram stain for diagnosing orthopaedic infections. *J Orthop Res* 24:1641–1649
21. García-Sierra N, Lacombe A, Prat C et al (2011) Pyrosequencing for rapid molecular detection of rifampin and isoniazid resistance in *Mycobacterium tuberculosis* strains and clinical specimens. *J Clin Microbiol* 49:3683–3686
22. Helgason E, Okstad OA, Caugant DA et al (2000) *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*—one species on the basis of genetic evidence. *Appl Environ Microbiol* 66:2627–2630
23. Ash C, Farrow JA, Dorsch M et al (1991) Comparative analysis of *Bacillus anthracis*, *Bacillus cereus*, and related species on the basis of reverse transcriptase sequencing of 16S rRNA. *Int J Syst Bacteriol* 41:343–346
24. Priest FG, Goodfellow M, Todd C (1988) A numerical classification of the genus *Bacillus*. *J Gen Microbiol* 134:1847–1882
25. Harrell LJ, Andersen GL, Wilson KH (1995) Genetic variability of *Bacillus anthracis* and related species. *J Clin Microbiol* 33:1847–1850
26. Jernigan DB, Raghunathan PL, Bell BP et al (2002) Investigation of bioterrorism-related anthrax, United States, 2001: epidemiologic findings. *Emerg Infect Dis* 8:1019–1028
27. Pearson T, Busch JD, Ravel J et al (2004) Phylogenetic discovery bias in *Bacillus anthracis* using single-nucleotide polymorphisms from whole-genome sequencing. *Proc Natl Acad Sci U S A* 101:13536–13541
28. Van Ert MN, Hofstadler SA, Jiang Y et al (2004) Mass spectrometry provides accurate characterization of two genetic marker types in *Bacillus anthracis*. *Biotechniques* 37:642–648
29. Hurtle W, Bode E, Kulesh DA et al (2004) Detection of the *Bacillus anthracis* *gyrA* gene by using a minor groove binder probe. *J Clin Microbiol* 42:179–185
30. Hill KK, Ticknor LO, Okinaka RT et al (2004) Fluorescent amplified fragment length polymorphism analysis of *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis* isolates. *Appl Environ Microbiol* 70:1068–1080
31. Ellerbrok H, Nattermann H, Ozel M et al (2002) Rapid and sensitive identification of pathogenic and apathogenic *Bacillus anthracis* by real-time PCR. *FEMS Microbiol Lett* 214:51–59
32. Qi Y, Patra G, Liang X et al (2001) Utilization of the *rpoB* gene as a specific chromosomal marker for real-time PCR detection of *Bacillus anthracis*. *Appl Environ Microbiol* 67:3720–3727
33. Joung KB, Cote JC (2002) Evaluation of ribosomal RNA gene restriction patterns for the classification of *Bacillus* species and related genera. *J Appl Microbiol* 92:97–108
34. Yamazaki K, Okubo T, Inoue N et al (1997) Randomly amplified polymorphic DNA (RAPD) for rapid identification of spoilage bacterium *Alicyclobacillus acidoterrestris*. *Biosci Biotechnol Biochem* 61:1016–1019
35. Goto K, Omura T, Hara Y et al (2000) Application of the partial 16S rDNA sequence as an index for rapid identification of species in the genus *Bacillus*. *J Gen Appl Microbiol* 46:1–8
36. Turenne CY, Tschetter L, Wolfe J et al (2001) Necessity of quality-controlled 16S rRNA gene sequence databases: identifying nontuberculous *Mycobacterium* species. *J Clin Microbiol* 39:3637–3648
37. Amoako KK, Shields MJ, Goji N et al (2012) Rapid detection and identification of *Yersinia pestis* from food using immunomagnetic separation and pyrosequencing. *J Pathog* 2012: 781652
38. Amoako KK, Janzen TW, Shields MJ et al (2013) Rapid detection and identification of *Bacillus anthracis* in food using pyrosequencing technology. *Int J Food Microbiol* 165:319–325
39. Loveless BM, Yermakova A, Christensen DR et al (2010) Identification of ciprofloxacin resistance by simpleprobe, high resolution melt and pyrosequencing nucleic acid analysis in biothreat agents: *Bacillus anthracis*, *Yersinia pestis* and *Francisella tularensis*. *Mol Cell Probes* 24:154–160
40. Wahab T, Hjalmarsson S, Wollin R et al (2005) Pyrosequencing *Bacillus anthracis*. *Emerg Infect Dis* 11:1527–1531
41. Ahmod NZ, Gupta RS, Shah HN (2011) Identification of a *Bacillus anthracis* specific indel in the *yeaC* gene and development of a rapid pyrosequencing assay for distinguishing *B. anthracis* from the *B. cereus* group. *J Microbiol Methods* 87:278–285
42. Janzen Timothy, Michael Shields, Noriko Goji, Matthew Thomas, Kristen Hahn and Amoako K (2015) Rapid detection method for *B. anthracis* using a combination of multiplexed real-time PCR and pyrosequencing and its application for food biodefense. *J Food Prot* 78:355–361
43. Shields MJ, Hahn KR, Janzen TW et al (2012) Immunomagnetic capture of *Bacillus anthracis* spores from food. *J Food Prot* 75:1243–1248
44. Amoako KK, Goji N, Macmillan T et al (2010) Development of multitarget real-time PCR for the rapid, specific, and sensitive detection of *Yersinia pestis* in milk and ground beef. *J Food Prot* 73:18–25

Free ebooks ==> www.Ebook777.com

Part V

Forensics

Free ebooks ==> www.Ebook777.com

Chapter 26

Forensic Analysis of Mitochondrial and Autosomal Markers Using Pyrosequencing®

Magdalena M. Buś, Hanna Edlund, and Marie Allen

Abstract

Forensic casework analyses often face challenges, such as limited genetic material with or without fragmentation and damage. To compensate for low amounts and degradation, shorter amplicons are often applied in the analysis. Also, a change of markers might be necessary using mitochondrial instead of autosomal markers. In addition, forensic research often involves analysis of large number of samples for marker evaluation and population-database compilation. Therefore, a flexible, robust but also rapid method for the detection of variation is highly useful. Pyrosequencing® is a rapid, reliable, easy-to-use method for sequence analysis. The method is well suited for rapid forensic analysis of a few targets or analysis of a single target in many samples. It allows sequencing of very short amplicons, which facilitates analysis of degraded DNA. Here we present the use of Pyrosequencing, a robust method for sensitive forensic analysis of mitochondrial DNA, autosomal STRs, and Y-chromosome STRs and SNPs.

Key words Pyrosequencing®, Forensic casework, Mitochondrial DNA, Autosomal DNA, STRs, SNPs

1 Introduction

Samples collected at a crime scene often have very limited amounts of DNA that are insufficient for routine forensic analysis. Common sources of DNA in forensic investigations are epithelial cells, hair, stains from saliva, blood, semen, or other body fluids found on various surfaces and items. Unfortunately, biological evidence is susceptible to several degradation factors, which makes it important to use optimal methods for preservation, DNA extraction, and further molecular analysis. Most forensic laboratories use fragment-size separation of short tandem repeats (STRs) for their routine analysis. The high power of discrimination for a panel of markers and the ability to perform a multiplex analysis provide a valuable identification tool for individuals. Other types of widely used markers in forensic investigations are Y-chromosome STRs or single-nucleotide polymorphisms (SNPs). These markers are commonly used to analyze and interpret mixtures from several individuals in

sexual assault cases. Furthermore, it is an advantage using autosomal SNPs for analysis of degraded samples, as the target size can be kept short. Also, markers on the sex chromosomes, such as variants in the amelogenin gene, are used for sex determination [1].

When the DNA in the forensic samples is severely degraded and therefore insufficient for standard STR analysis, mitochondrial DNA (mtDNA) can provide an alternative genetic source for typing. The power of discrimination in an analysis based on mtDNA is significantly lower in comparison with unlinked autosomal markers. However, due to the considerably greater copy numbers, analysis of mtDNA is very often successful when challenging samples are examined. Typing of degraded DNA using nuclear (nDNA) or mtDNA requires a relatively short length of the amplified fragments for successful results. In the routinely performed mtDNA analysis, sequences of the hypervariable regions (HVI and HVII) are determined using the Sanger sequencing method [2, 3]. The method is considered the gold standard for sequencing, although the technology has some limitations when challenging samples are involved in the analysis. One limitation is the fact that Sanger sequencing generates unreadable sequence signals roughly 20–30 bases downstream of the primer. As reads for both strands are preferred and most amplicons in the forensic DNA analysis do not exceed 200 bp, there is a loss of an average of 20–30 % of genetic information, with the assumption that the rest of sequence is of good quality.

Forensic samples frequently show a mixture of several mtDNA profiles from an unknown number of individuals. Mixtures can also be seen in the sequence data as a result of heteroplasmy or contamination and will make the interpretation of the sequencing results even more complex [4, 5]. Although variation within mtDNA can be detected and visualized using Sanger sequencing, it is not possible to get an accurate quantification of the different components of a mixture. However, due to the fact that the amount of released pyrophosphate in a Pyrosequencing reaction is in proportion to the quantity of the template, a quantification of mixtures can be performed using the Pyrosequencing technology. This advantage is utilized in a quantification assay for mtDNA mixtures with the ability to interpret major and minor components in forensic mixtures [6]. Moreover, assays for sequence analysis of the HVI, HVII, and coding region have been developed based on Pyrosequencing. It was first demonstrated that the Pyrosequencing technology provides accurate sequencing results for determination of variable sites in the D-loop from many types of forensic evidence samples [7]. Later an additional assay for analysis of polymorphisms in the coding region was developed for an increase in the discrimination power of mtDNA analysis [8].

Forensic crime scene samples constitute an invaluable source of genetic information, and the most informative results are given

when the assay has high discrimination power. An easy switch between different types of markers will make the best choice for further analysis for challenging samples easier (e.g., mtDNA or nDNA, longer or shorter fragments). Therefore, we have developed several robust, flexible, and rapid typing systems based on Pyrosequencing for analysis of nuclear markers like autosomal STRs, Y-STRs, Y-SNPs, and a deletion in the amelogenin gene [9, 10]. These assays are optimal for samples with a limited degree of degradation and DNA amounts that are sufficient for a nuclear DNA analysis. Although Pyrosequencing was first developed for analysis of SNP variation or sequencing of short DNA fragments, STR analysis is possible, though the interpretation is more complex. The allele determination is based on sequence composition, unlike fragment-size differences as in routine analysis. To facilitate interpretation of STRs, we have developed a method based on using a direct dispensation order and a termination-recognition base (TRB). The TRB is the first base occurring downstream that is not part of the repeat unit, and the complementary base is not dispensed in the reaction. When the reaction reaches the TRB for the shortest allele of the STR marker in heterozygous samples, half the signal will be displayed in the continued Pyrogram®. The second allele is determined when the reaction reaches the TRB for the longer allele. Y-STR interpretation is less complex, and the number of repeats in the sequence prior to the TRB corresponds to the specific allele. Although the routine capillary electrophoresis-based method for STR analysis is robust and sensitive, it requires fairly long amplicons to avoid size overlaps. In contrast, as the Pyrosequencing assay is not based on size separation all targets can be kept as short as possible according to the repeat sequence. Even shorter targets are allowed when SNP analysis is performed, and the interpretation is unproblematic as long as the SNP does not involve a homopolymeric region. One drawback in using Pyrosequencing, however, is the limited possibility of multiplexing a large number of targets in the same reaction.

In some investigations the knowledge of the sex of the donor of a sample is highly important. Therefore, we developed a Pyrosequencing approach based on determination of an indel variant (insertion/deletion) in the amelogenin gene (AMEL). The gene is located on both Y (AMELY) and X (AMELX) chromosomes, and the X chromosome has a 6 bp deletion that is commonly used for sex determination in forensic analysis. A homozygous genotype with 6 bp deletion on both chromosomes will thus reflect a female, whereas a heterozygous pattern identifies a male sex [11].

Here we present several protocols for a sensitive, robust, and accurate forensic analysis using the Pyrosequencing technology. Although mtDNA analysis has been discussed here, further focus is on protocols for sensitive analysis of autosomal markers such as nSTRs, Y-STRs, Y-SNPs, and AMEL(X/Y).

2 Materials

2.1 Primer Design

For example Primer Express® Software package (Life Technologies) or Primer3 [12] (*see* **Note 1**).

2.2 PCR Reaction

1. PCR strips and caps, e.g., MicroAmp® 8-Cap Strip and MicroAmp® 8-Tube Strip (Life Technologies).
2. 0.6 or 1.5 mL tubes (depending on the number of analyzed samples), e.g., boil-proof, maximum recovery, clear (Axygen).
3. Primer pairs with one unlabeled primer and a second with a 5' biotin label. For more suggestions *see* **Note 1**.
4. AmpliTaq Gold® with GeneAmp® containing AmpliTaqGold® DNA Polymerase (5 U/μL) (*see* **Notes 2** and **3**), 10× PCR Gold Buffer without MgCl₂, and 25 mM MgCl₂ Solution (Life Technologies).
5. dNTP set.
6. Glycerol.
7. BSA (*see* **Notes 3** and **4**).

2.3 Gel Electrophoresis

1. Agarose, e.g., SeaKem® LE Agarose (Lonza) or equivalent.
2. SYBR®Safe DNA gel stain (Invitrogen, Life Technologies) or equivalent dye for DNA detection in gels.
3. 1× TBE buffer.

2.4 Template Preparation

1. PCR products.
2. Sequencing primer(s) (*see* **Note 1**).
3. PCR strips, and cap strips.
4. 0.6 or 1.5 mL tubes (depending on the number of analyzed samples).
5. PyroMark®Q24 Plate (Qiagen).
6. PyroMark®Q24 Vacuum Prep Workstation (Qiagen).
7. PyroMark®Q24 Cartridge (Qiagen) (*see* **Note 5**).
8. PyroMark® Gold Q24 Reagents (Qiagen) containing enzyme mixture, substrate mixture, and nucleotides dATPαS, dGTP, dCTP, dTTP (*see* **Note 6**).
9. Streptavidin Sepharose® High Performance (e.g., GE Healthcare Bio-Sciences AB) (*see* **Note 6**).
10. PyroMark® Denaturation Solution 500 mL (Qiagen).
11. PyroMark® Annealing Buffer 250 mL (Qiagen).
12. PyroMark® 10× Wash Buffer 200 mL (Qiagen).
13. PyroMark® Binding Buffer 200 mL (Qiagen).

14. Ethanol.
15. High-purity water (recommendation for PyroMark™Q24 manufacturer is Milli-Q 18.2 MΩ cm or equivalent).

2.5 Pyrosequencing

1. PyroMark™Q24 instrument (Qiagen).
2. USB memory stick.

2.6 Data Analysis

1. PyroMark™Q24 software.

3 Methods

3.1 Autosomal STR Typing Protocol

Markers: TPOX, CSF1PO, D5S818, D16S539, D8S1179, PentaE, D13S317, D7S820, TH01, D3S1358.

A singleplex reaction mix: 1×PCRTaq Gold Buffer, 0.2 mM dNTPs, 2.5 mM MgCl, 4 % glycerol, 0.2 μM of each primer (Table 1), 0.025 U AmpliTaqGold® DNA Polymerase, and up to 20 ng genomic DNA, in a total volume of 35 μL. Thermal cycling is performed as described in Table 2.

A duplex reaction can be performed for markers, TPOX/CSF1PO, D5S818/ D16S539, D8S1179/PentaE, D13S317/D7S820, and TH01/D3S1358, using the same protocol as for single-reaction amplification, but with a total volume of 50 μL.

3.2 Typing of Y-STRs, Y-SNPs, and AMEL(X/Y)

Markers:

1. Y-STRs: DYS19, DYS389I, DYS389IIA, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS438.
2. Y-SNPs: M26, M34, M35, M42, M45, M55, M56, M67, M89, M96, M119, M137, M160, M170, M173.
3. AMEL(X/Y)—106 bp (XX) and 112 bp (XY).

A singleplex reaction for amplification of Y-STR, Y-SNP, and AMEL(X/Y) markers contains 1×PCR Gold Buffer, 0.2 mM dNTPs, 1.5 mM MgCl₂, 4 % glycerol, 0.2 μM of each primer (Table 1), 1 U AmpliTaqGold DNA Polymerase, and 1–10 ng genomic DNA, in a total volume of 35 μL. The details for the thermal cycling conditions are presented in Table 3.

Amplification of forensic low-template samples (all markers): 1×PCR Gold Buffer, 0.2 mM dNTPs, 1.5 mM MgCl, 0.2 μM of each primer (Table 1), 10 % glycerol, 0.16 mg/mL BSA, 5 U AmpliTaqGold DNA Polymerase, in a total volume of 30 μL, with the same cycling conditions as above.

3.3 Pyrosequencing PyroMark™Q24 Device Protocol

1. Allow all the chemical reagents to reach room temperature (*see Note 5*).
2. Start the PyroMark™Q24 instrument and a computer with the Q24 software. Create a new run method, save the new run

Table 1
PCR and sequencing primers for Pyrosequencing

Primer name	Sequence 5' → 3'	Sequencing primer	Pyrosequencing dispensation
<i>nSTR</i>			
TPOX	F-AGAACAGGCACCTTAGGAA Bio-R-AGCGTTTATTTGCCAA	X	CTCACTG[ATG] ₁₆
D3S1358	F-CCAACTGGGTGACAGAGCA R-ACTCATGAAATCAACAGAGGCTT	X	GCAATGTA[TCTA][TCTG] ₃ [TCTA] ₁₇ T
D5S818	F-CATTGTATCTTTATCTGTATCCCTTA Bio-R-CCTCTTTGGTATCCTTCTGTAATA FSEQ-GTATCTTATCTGTATCCTTATTT	X	ATACTCT [ATAC] ₂₁
D7S820	F-GATAGAACACTTGTGCATAGTTTAGAA Bio-R-GCACCAAAATTTGGTAATTAAA	X	CTGACTAC [GATA] ₁₆
D8S1179 ^a	F-GTATTTTCATGTGTACATTCGTATCTA Bio-R-GCGTGAATATGCCTTAATTTA	X	[TCTA][TCTG] ₂ [TCTA] ₁₇
D13S317	F-CATCTAACGCCCTATCTGTATTTTACAA Bio-R-GCCCCAAAAAGACAGACAGAA	X	ATACA [TATC] ₂₁
D16S539	F-GGAGCAAAACAAAGGCAGA R-CCATCTCTGTTTGTCTTTCAAT	X	GA[TATC] ₁₇
TH01	F-CATTGGCCTGTTCTCCCTTGA Bio-R-GCAGGTCACAGGGAACACAGA	X	[CAT] ₁₇
CSF1PO	F-AGATATTAAACAGTAACTGCCCTCA Bio-R-CAGATACCTATCTCCTGGTGCA FSEQ-TTCATAGATAGAAAGATAGATGATT	X	[AGAT] ₁₈
Penta E	Bio-F-GGCGACTGAGCAAGACTCA R-GGTTATTAAATTGAGAAAACTCCTTACAA ^b RSEQ-GAGAAAACTCCTTACAATTT	X	[TCG] ₂₄

<i>γ-STR</i>			
DYS19	Bio-F-CTGAGTTTCTGTATAGTGTTTTAA R-CTGGGTTAAGGAGAGTGCA	X	[TCTA] ₂₁
DYS389 I	Bio-F-CCAACTCTCATCTGTATTATCTAIGT R-GATAGATTGATAGAGGGAGGA	X	[TAGA] ₁₄ [CAGA] ₆
DYS389 IIA ^c	F-CAACTCTCATCTGTATTATCTATGTGTGT Bio-R-GATAGATTGATAGAGGGAGGA	X	[TCTG] ₅ [TCTA] ₁₉
DYS389 II	Bio-F-GTATCCAACTCTCATCTGTATTATCTA R-GATAGATTGATAGAGGGAGGA	X	[TA] ₁₃ [CA] ₃ [TA] ₁₄ [CA] ₅ ^d
DYS390	F-GCAATTTGGTACCCCATAA Bio-R-GTATACTCAGAAAACAAGGGAAGA	X	[TG] ₈ [TA] ₁₆ [TG] [TA] ₁₀ ^e
DYS391	F-CATTCAATCATACACCCATATCTGT Bio-R-CCATAGAGGGATAGGTAGGCA	X	CTGTCTG [TCTA] ₁₈
DYS392	F-TGATTTCAAAGTGTTTGTATTTTAA R-ACCAATCCCATTCCTTAGTAA	X	[TA] ₂₂
DYS393	F-GTGGTCTTCTACTTGTGTCAATAC Bio-R-AACTCAAGTCCAAAAAATGAGG	X	C[AGAT] ₂₀
DYS438	F-TTGAACGGTAAACAGTATATTTTC Bio-R-AGAGTGAAACTCCATTTCAAATAGA	X	[TC] ₁₆
<i>γ-SNP</i>			
M26	F-GATTCAAATTTTCTGAATTAGAA Bio-R-GAATTCATAGGCCAATCAGT	X	
M34	F-GTAGTCACAGTGTTTCTCATGTAA Bio-R-TTTGCAGACACACCACATGT	X	CTGCTGCTGACAGAG/TGC
M35	F-CAAATTTTCTTTGGGAC Bio-R-AGACTTTCGGAGTCTCTGCC	X	

(continued)

Table 1
(continued)

Primer name	Sequence 5' → 3'	Sequencing primer	Pyrosequencing dispensation
M42	F-GCAACTTAATCAGATTTTAGGACA Bio-R-GTCACCCAGCTCTCTTTTCA	X	GCAAGCT/AGC
M45	F-AAATTGGCAGTGAAAAATTATA Bio-R-GGACCTCAGAAGGAGCTT	X	CGATA/GTCA
M55	F-GTGAAAAAGGCCCTGG Bio-R-TCTCTTCTCATTCITTTAACTCTTA	X	GATGTAGTAT/CGA
M56	F-CCAGAAACTGAAGTACAAATGCA Bio-R-CTGACCATACTTATCATCACTTAGCC	X	GATGAGATACGA/TCG
M67	F-GTACTCAAAATATGTGTAATTCAAAA Bio-R-GTTCTGTGGACCCCTCTAT	X	GACA/TATATAAG
M89	F-CTGGATTCAGCTCTCTTCCTAA Bio-R-CCAGGATCACCAGCAAAG	X	CGTATGTACAATCT/CGA
M96	F-CATATAATTTGCCATAGTTTTT Bio-R-GCTGTGATGTGTAACCTTGGAA	X	GATATATACTGAGTGTATC/GCT
M119	F-GGGTTATTCCAATTCAGCATA Bio-R-TTGGGGAGACAGATAATTCTG	X	GCAGCT/GCA

M137	F-GTTGAAGAAAACTTTATCTATTG Bio-R-GGTTTAATAATACTGAGTTTGTGC	X	CGATTGCTCTA/CGT
M160	F-GTTCACAAAACTTCTGAATTACAA Bio-R-CCCTCTGAACGCTAACCACAA	X	CGTATACATACTCATC/AGC
M170	F-ATTCTGTGCATTATACAAATTACTATT Bio-R-CAGGTCCCTCAITTTTACAGTGA	X	GTATACGAATCATGTC/AGTT
M173	F-TCTTACAATTCAAGGGCAITTTA Bio-R-CTGGCTTATCAITTTCTGAATATTAA	X	CGAC/ACATGT
Amelogenin	F-CCCTGGGCTCTGTAAAGAATAGT Bio-R-ATCAGAGCTTAAACTGGGAAGCTG	X	GTGTGGATCTCATCAGATA GTGTCTCAGTGTCT

^aForward primer covers the first TCTA repeat unit

^bThe Penta E R PCR primer can also be used as sequencing primer

^c(TCTG)_n(TCTA)_p of the DYS389 II repeat (TCTG)_n(TCTA)_p(TCTA)_r(TCTA)_r

^dMixture of dNTPs (A/G) in the Pyrosequencing reaction (A dispensed)

^eMixture of dNTPs (T/C) in the Pyrosequencing reaction (T dispensed)

Table 2
Thermal cycling conditions for amplification of nSTRs

Markers	Cycling conditions	Annealing temperature (°C)
TPOX, CSF1PO, D5S818, D16S539, D8S1179, PentaE, D13S317, D7S820	10 min at 95 °C, 45 cycles of 30s at 95 °C, 30s at annealing temperature, 30s at 72 °C, and a final extension step of 7 min at 72 °C	53
THOI D3S1358		60

Table 3
Thermal cycling conditions for amplification of Y-STRs, Y-SNPs, and AMEL(X/Y) markers

Markers	Cycling conditions	Annealing temperature (°C)
Amelogenin, DYS19, DYS389-I, DYS390, DYS391, DYS392, DYS393, DYS438, M26, M42, M45, M55, M56, M67, M89, M96, M119, M137, M160, M170	10 min at 95 °C, 45 cycles of 30 s at 95 °C, 45 s at annealing temperature, 60 s at 72 °C, and a final extension step of 7 min at 72 °C	55
M34, M35, M173		58
DYS389-II DYS389-IIA		60

information to a USB memory stick, and print the reagent volume information sheet.

3. Set a heating block at 80 °C.
4. Prepare a PyroMark™Q24 Vacuum Prep Workstation by filling the five trays in the workstation with freshly prepared solutions: approximately 50 mL 70 % ethanol, approximately 40 mL denaturation solution, approximately 50 mL 1× washing buffer, and high-purity water—one tray with approximately 50 mL and a second tray with 70 mL. Mark the specific position of the different solutions properly. The Vacuum Prep Workstation tool is now ready to use for generation of single-stranded DNA templates from the PCR products.
5. Prepare:
 - (a) A PyroMark™Q24 Plate.
 - (b) PCR 8 tube strip(s).
 - (c) 0.6 or 1.5 mL tubes.
 - (d) Primer(s) for sequencing from a freezer.
 - (e) PCR products.

Table 4
Details for preparation of Sepharose bead and primer mixes

Sample number	Reaction Sepharose bead mix (point 6 in the Pyrosequencing Q24 device protocol)			Primer (10 μ M) solution mix calculation (point 9 in the Pyrosequencing Q24 device protocol)	
	Streptavidin Sepharose beads (μ L)	Binding buffer (μ L)	dH ₂ O (μ L)	Annealing buffer (μ L)	Primer (μ L)
1	2	40	18	24.25	0.75
10	20	400	180	242.50	7.50
25	50	1,000	450	606.25	18.75

6. In a 0.6 or 1.5 mL tube prepare a Sepharose Bead mix (follow instructions in Table 4) (*see* **Note 5**).
7. In PCR tube strip(s) mix 60 μ L of Sepharose Bead mix with 20 μ L of PCR product.
8. Shake the strip(s) with the beads for 10 min at room temperature (speed approximately 1,000 rpm).
9. In a 0.6 or 1.5 mL tube, prepare the sequencing primer mix (follow the instructions in Table 4).
10. Add 25 μ L of sequencing primer mix to each well in the PyroMark Q24 Plate (Table 4).
11. Prepare a PyroMark Q24 Cartridge according to information printed in the reagent volume information sheet.
12. Stop the shaker and place both the PyroMark Q24 Plate and PCR tubes in the PyroMark Q24 Vacuum Prep Workstation in appropriate positions.
13. Switch on the vacuum pump. Apply the handheld vacuum tool into the strip(s) containing the immobilized templates and open the switch. Carefully put the filter probes into the PCR strip(s) and shake gently from side to side/up and down for 15–20 s. Make sure that all of the solution is aspirated from the wells and that the beads have been captured onto the filter probe tips.
14. Transfer the tool to the well containing 70 % ethanol and hold for 5 s.
15. Transfer the tool to the well containing denaturation solution and hold for 5 s.
16. Transfer the tool to the well containing washing buffer and hold for 10 s.

17. Hold the tool over the PyroMark Q24 Plate and switch off the vacuum pump and the tool before the filter probes reach the sequencing primer solution.
18. Release the beads in the plate containing sequencing primer by shaking the tool gently from side to side for approximately 20 s.
19. Transfer the tool to the well containing high-purity water and agitate for 10 s.
20. Transfer the tool to the next well containing high-purity water and agitate for 10 s.
21. Place the tool in the parking position.
22. Heat the PyroMark Q24 Plate containing the samples at 80 °C for 2 min and move to room temperature.
23. Place the PyroMark Q24 Cartridge with reagents and the PyroMark Q24 Plate in the instrument and close the lid.
24. Insert the USB memory stick containing the run file, select the appropriate run name, and press the run button.

3.4 Genotyping and Interpretation of Pyrograms

The dispensation of nucleotides in a Pyrosequencing reaction can be either cyclic or directed. In cyclic dispensation, all four nucleotides are repeatedly added (ACGTACGT, etc.). This option is recommended for de novo sequencing or for targets with a high degree of variation. A sequence-directed dispensation order is pre-programmed in accordance with the expected sequence and variants. This option was developed and used for autosomal and Y-chromosomal STR analysis, Y-SNP typing, and sex determination using AMEL(X/Y). Dispensations orders are shown in Table 1.

3.4.1 STRs

To allow interpretation of different alleles in analysis of Y-STRs and autosomal STRs, a termination recognition base (TRB) is used. A TRB is a nucleotide that is not part of the repeat unit. As the TRB is not dispensed in the sequencing reaction, it results in a termination of the reaction when the corresponding base occurs in the sequence. In heterozygous genotypes, termination of sequencing of the shorter allele will result in a decrease of the signal by half, whereas no signal reduction is observed in homozygous genotypes. The Pyrosequencing reaction then proceeds until the longer allele reaches the TRB and the reaction comes to a complete stop. In addition to counting the number of repeats, the unique pattern observed in the Pyrogram after the TRB will aid in the interpretation. Thus, for each heterozygous or homozygous genotype, different sequence patterns are observed in the resulting Pyrograms depending on both sequence composition and length to the TRB (Fig. 1).

3.4.2 Y-STRs

Due to the haploid nature of Y-STR loci, only one allele per locus is displayed for an individual, and the Pyrograms can be interpreted in a straightforward way by calculating the number of repeats

Table 5
Redesigned sequencing primers for multiplex Pyrosequencing of Y-SNP markers

Y-SNP	Sequencing primer
M26	TTTTCTGAATTAGAATTA
M34	GCTTCCACCCAGGAG
M35	–
M42	AACTTAATCAGATTTAGGACACAAA
M45	AAATTGGCAGTGAAAAATTATA
M55	GCCCCTGGGATGGTTTAA
M56	TGCAATGGGAGGATTACGA
M67	ATATGTGTAATTCAAAAAAC
M89	–
M96	AATATTATACCTGAGTGTTT
M119	ATTCCAATTCAGCATACA
M137	TATTGAGATTTTGCTTTCC
M160	ACATACAACCTCAATTTTC
M170	ATTTACTTAAAAATCATTGT
M173	CAATTCAAGGGCATTTA

3.4.4 Y-SNPs

The PyroMark Q24 software has been developed for SNP analysis and is therefore well suited for Y-SNP interpretation. The quality of the resulting data is based on signal-to-noise ratio and peak heights, and the Pyrograms are scored as passed, checked, or failed. Interpretation of single Y-SNP markers is unproblematic, and multiplex Y-SNP analysis is also possible with a carefully designed system (Table 5). The sequencing result is evaluated by comparison between the predicted Pyrogram from the software and the actual signals in the Pyrogram (Fig. 4).

4 Notes

1. Since Pyrosequencing is able to yield analyzable data from the first base next to the annealed primer, primer design becomes more flexible than in Sanger sequencing. As challenging samples often are degraded into smaller fragments, a shorter target size will improve the success rate. Primers can be placed just before or even next to the target SNP or STR and therefore the entire target can be very short if the surrounding sequence

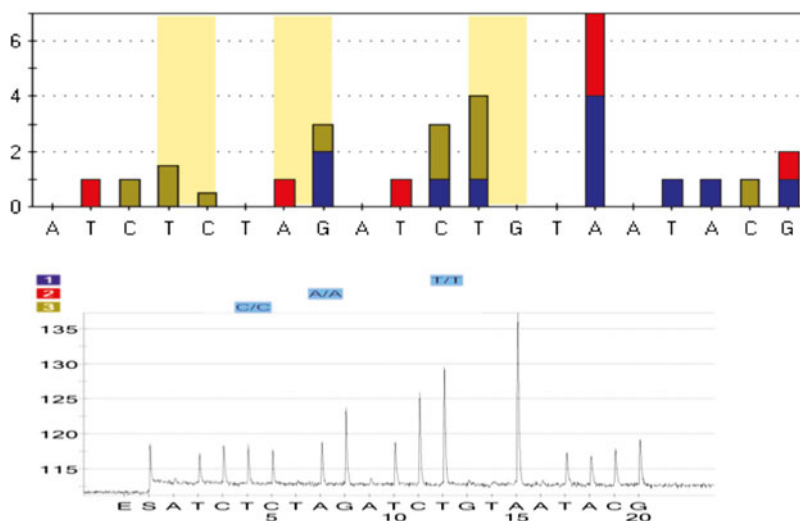


Fig. 4 Multiplex Pyrogram of the Y-SNP markers M145, M137, and M119. (a) The software-predicted Pyrogram. The *green* peaks represent the sequence for marker M145 (CTCGCTTTC), the *red* peaks are for the M137 marker (TATAAAG), and *blue* peaks represent M119 (GGCTAAAATAG). The *yellow* bars indicate SNP positions. (b) The actual Pyrogram demonstrating an M145C, M137A, and M119T haplotype

allows such a design. Nevertheless, a design with 5–10 nucleotides before and after the target fragment is recommended, as it makes the interpretation of results much easier to perform. The read length is a limitation in Pyrosequencing, as a maximum of 100 dispensations can be performed in a run. However, designs for shorter STRs or even several SNPs located in one fragment are possible within this range.

It is important to order one of the PCR primers biotinylated to allow preparation of ssDNA prior to sequencing. Usually, it is of no importance if it is the forward or reverse primer that is biotinylated. The unbiotinylated primer designed for PCR can be utilized as sequencing primer, although an internal primer will in most cases decrease the background signal. Primers should be checked for self-looping, primer-dimer, and mispriming as in all PCR design, and this is highly important in the design of multiplex assays. In general, a design for single Pyrosequencing reactions is straightforward, while in multiplex reactions primer-dimer structures or nonspecific amplification are easily formed, causing a high background. Therefore, a redesign of at least some of the primers is recommended if going from singleplex to multiplex systems, as demonstrated by the Y-SNP multiplex analysis described here (Table 5). Primers, especially biotin-labeled, should be HPLC purified because free biotin will compete with biotinylated PCR products for binding sites on the Sepharose beads.

2. To prevent contamination the PCR amplification should be set up in a laboratory with separate areas for DNA and reagents, which should be physically separated from the post-PCR area. Laboratory benches and equipment are optimally cleaned with 15 % NaOCl (bleach) and subjected to UV radiation for at least 1 h prior to use [13]. Protective gloves, lab coats, facemasks, and hair protection caps should be used in all pre-PCR steps. Moreover, disposable gloves must be changed frequently.
3. A large variety of biological material can be a source of DNA in forensic investigations (e.g., blood, saliva, bone, skin, semen, and body fluids). The samples can also vary in quality and quantity depending on the initial DNA copy numbers and the degree of degradation and damage. These factors in combination with a presence of inhibitors will cause differences in PCR efficiency between samples. The manufacturer of the PyroMark Q24 device recommends an input of 10 ng human genomic DNA per PCR reaction for standard Pyrosequencing using the PyroMark PCR Kit (Qiagen). However, based on our experience much smaller amounts of DNA can be used with maintained quality and readable Pyrograms. We recommend carrying out a sensitivity test for evaluation of the minimal amount of DNA required for each specific marker. Although the PCR and the following Pyrosequencing are very robust some PCR modifications are sometimes required. In the attached PCR protocols AmpliTaqGold DNA Polymerase was used; however, use of other types of PCR polymerases can be beneficial in some cases. In general, using a hot start polymerase is recommended, as it decreases the formation of primer-dimers and increases the sensitivity of the assay. Various types of inhibitors are common and might severely affect PCR efficiency. To improve analysis if inhibitors are present, further purification of samples or addition of BSA may be considered. Due to the nature of forensic samples, correct measurement of the quantity of DNA is crucial for good results. Therefore, we recommend a method based on real-time PCR or fluorescence, as these are more accurate than a spectrophotometric method. Real-time PCR, in contrast to other quantification assays, will indicate the presence of inhibitors by inefficient amplification curves.
4. Some additional tricks can be applied to reduce the background and thereby facilitate the interpretation of the Pyrograms. Both the negative PCR control and a PCR product without sequencing primer can help to monitor the Pyrosequencing background signal. When directed dispensation is used, single-control nucleotides (SCNs) that are not part of the expected sequence can be inserted into the dispensation order. This approach is especially useful in STR analysis

to provide a control for nonspecific peaks. The “control nucleotides” can be inserted before and after each STR repeat unit and will be very helpful in the interpretation.

The progressive process of a Pyrosequencing reaction will result in depletion of nucleotides and a decrease in the enzyme efficiency. The peak heights will therefore decrease with the length of a dispensed sequence. For example, a peak for AAA at the beginning of the run will be higher than the same poly-A sequence 50 nucleotides later. For easier interpretation of homopolymeric regions, especially in the distant part of a sequence, the addition of SCNs will be helpful.

The sequencing of problematic templates with for example homopolymeric stretches or GC-rich templates can be improved by using certain additives as single-stranded DNA-binding protein (SSB). SSB will improve the signal intensity and increase the reaction efficiency, especially for long reads, by a stabilizing effect and prevention of the formation of secondary structures in ssDNA. If BSA is used as a component of the PCR reaction to counteract inhibition, its concentration must be accurately optimized. Excessively high concentration inhibits PCR reaction or causes background noise in Pyrograms.

5. All reagents must be changed, checked for expiration date, or prepared fresh before a new analysis is performed according to instructions in the protocols. During storage at +4 °C, streptavidin Sepharose beads will sediment and split the solution into two visible phases. Therefore, it is very important to intensively agitate a bottle with the reagent into a homogeneous solution before use. All sample preparation solutions (annealing, binding, washing buffer, and denaturation solution), unopened PyroMark Q24 Gold Reagents, and dNTPs should be stored at +4 °C. Resolved enzyme and substrate are stable at 2–8 °C for several days. Prepared enzyme and substrate mixtures can be stored at –20 °C for a longer time (several weeks), but repeated freezing and thawing should be avoided.
6. It is highly important to take care of the reagent cartridge for the Pyrosequencing instrument. Since small particles and reagents remain that can block the needles of the PyroMark Q24 Cartridge, it must be cleaned carefully, using high-purity water. It is crucial to check that all positions dedicated for specific reagents are passable by water. Especially the enzyme and substrate (E and S) needles are commonly blocked and can thereby prevent dispensation during the run. After a run the compartments should be rinsed at least four times with h326206 water. Test whether all the needles are clean by filling the cartridge with high-purity water and carefully pressing on the top of every reagent position. A thin jet of water should squirt out from the tip of the needle. If the needle is blocked even after

several washes, leave the cartridge filled with water for at least 1 h. Blockage occurs frequently when the cartridge is left in the instrument overnight without cleaning. It is also recommended that the cartridge should be changed after 30 runs.

Acknowledgements

M.A. received financial support from the Kjell and Märta Beijer Foundation, Vinnova (the Swedish Governmental Agency for Innovation Systems), and Brottsoffermyndigheten (the Swedish Crime Victim Compensation and Support Authority).

References

1. Gibbon V, Paximadis M, Strkalj G et al (2009) Novel methods of molecular sex identification from skeletal tissue using the amelogenin gene. *Forensic Sci Int Genet* 3:74–79
2. Budowle B, Wilson MR, DiZinno JA et al (1999) Mitochondrial DNA regions HVI and HVII population data. *Forensic Sci Int* 103:23–35
3. Lembring M, van Oven M, Montelius M et al (2013) Mitochondrial DNA analysis of Swedish population samples. *Int J Legal Med* 127:1097–1099
4. Calloway CD, Reynolds RL, Herrin GL Jr et al (2000) The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age. *Am J Hum Genet* 66:1384–1397
5. Cowell RG, Lauritzen SL, Mortera J (2007) Identification and separation of DNA mixtures using peak area information. *Forensic Sci Int* 166:28–34
6. Andréasson H, Nilsson M, Budowle B et al (2006) Quantification of mtDNA mixtures in forensic evidence material using pyrosequencing. *Int J Legal Med* 120:383–390
7. Andréasson H, Asp A, Alderborn A et al (2002) Mitochondrial sequence analysis for forensic identification using pyrosequencing technology. *Biotechniques* 32:124–133
8. Andréasson H, Nilsson M, Styman H et al (2007) Forensic mitochondrial coding region analysis for increased discrimination using pyrosequencing technology. *Forensic Sci Int Genet* 1:35–43
9. Edlund H, Allen M (2009) Y chromosomal STR analysis using Pyrosequencing technology. *Forensic Sci Int Genet* 3:119–124
10. Divne AM, Edlund H, Allen M (2010) Forensic analysis of autosomal STR markers using Pyrosequencing. *Forensic Sci Int Genet* 4:122–129
11. Kjellström A, Edlund H, Lembring M et al (2012) An analysis of the alleged skeletal remains of Carin Goring. *PLoS One* 7: e44366
12. Untergasser A, Cutcutache I, Koressaar T et al (2012) Primer3 new capabilities and interfaces. *Nucleic Acids Res* 40:e115
13. Buś MM, Allen M (2014) Collecting and preserving biological samples from challenging environments for DNA analysis. *Biopreserv Biobank* 12:17–22

Chapter 27

Tissue-Specific DNA Methylation Patterns in Forensic Samples Detected by Pyrosequencing®

Joana Antunes, Kuppareddi Balamurugan, George Duncan,
and Bruce McCord

Abstract

In certain circumstances the outcome of a trial may hinge on the ability of a forensic laboratory to determine the identity of biological stains present at crime scenes. An example of such a situation would be the detection of blood, saliva, vaginal fluid, or other body fluid in a specific location whereby its presence would reinforce the victim's or suspect's version of the events that happened during the commission of a crime. However, current serological methods used for identifying body fluids may lack the sensitivity and specificity to identify these fluids, particularly for trace levels. New procedures using proteomic methods and RNA-based gene expression show promise in addressing this issue; however, concerns about stability and relative levels of gene expression remain. An alternative approach is to utilize patterns of epigenetic DNA methylation. DNA methylation is an epigenetic mechanism that regulates the specificity of genes being expressed or silenced in cells. Regions in the human genome referred to as tissue-specific differentially methylated regions account for unique patterns of DNA methylation that are specific for each cell type. This chapter addresses the application of bisulfite-modified PCR combined with Pyrosequencing® to detect tissue-specific DNA methylation patterns and perform trace serological analysis. The quantitative nature and precision available with Pyrosequencing presents major advantages in these studies as it permits detection of and contrast between cells with differential levels of methylation. The procedure can be applied to a variety of biological fluids which may be present at crime scenes.

Key words DNA methylation, Pyrosequencing®, Bisulfite conversion, Body fluid identification, Forensics, Tissue-specific differentially methylated regions

1 Introduction

It is common practice to screen evidence in forensic casework for the presence of body fluids prior to DNA processing. This is a critical step in the evaluation of the probative nature of the evidence. An example of the use of this practice is in the collection of evidence with respect to suspected sexual abuse of minors, where the presence of DNA from skin epithelial cells may indicate innocent contact but the presence of DNA from intimate body fluids may indicate sexual abuse. Serological methods currently available in

forensic laboratories rely on the detection of specific proteins and are generally presumptive in nature. These tests are commonly based on immunochromatographic methods [1]. Because these tests depend on colorimetric reactions following antibody-based detection, they also present varying degrees of stability and specificity. For example, saliva is determined through the detection of α -amylase. However this enzyme may become degraded with time and can be present in other body fluids, such as urine [2]. Similarly, confirmation of the detection of semen can be achieved by microscopic observation of spermatozoa; however some males produce little or no spermatozoa.

To overcome these limitations, several new approaches are being developed including advanced proteomics methods involving tandem mass spectrometry, detection of tissue-specific RNA, and use of epigenetic techniques. Unfortunately, the detection of biomarkers by mass spectrometry requires expensive equipment and expertise.

A useful and sensitive alternative is the detection of mRNA in body fluids [2–6]. However, there are several limitations in the use of this technique. For example the quantity of biomarker mRNA detected must be normalized with respect to a housekeeping gene [7]. In addition, DNaseI is commonly used to eliminate any residual DNA left in the sample [4–6, 8, 9]. The use of this enzyme can create problems in DNA analysis laboratories through its release and accidental digestion of evidence.

The above problems can be avoided by utilizing epigenetic DNA markers. DNA is more stable than RNA or proteins, and is a logical choice for the discrimination of body fluids as it is the basis for current protocols used in forensic laboratories. Epigenetic processes including DNA methylation regulate many specific functions in cells. These include regions of the genome that control cell differentiation and others that produce transient changes dependent on outside stress [10, 11]. For example Eckhardt et al. [12] demonstrated that the mean methylation difference in certain regions of the genome is lower when comparing males with females (0.1 %) and different age groups (0.275 %) than when comparing different body tissues (7.1 %). On the other hand Li and colleagues [13] demonstrated that the environment influences the methylation status of some genes by comparing 3616 CpG sites in 22 groups of monozygotic twins.

DNA methylation studies using microarrays have identified tissue-specific differentially methylated regions (tDMRs) that seem to be associated with specific functions of each cell [14]. These regions have been found within or associated with gene promoters and CpG islands. Recent reports indicate that tDMRs also occur in alternative transcription start sites and regions outside CpG islands, namely locations that flank the CpG islands [15]. These data indicate the potential application of epigenetic methylation in forensic biology.

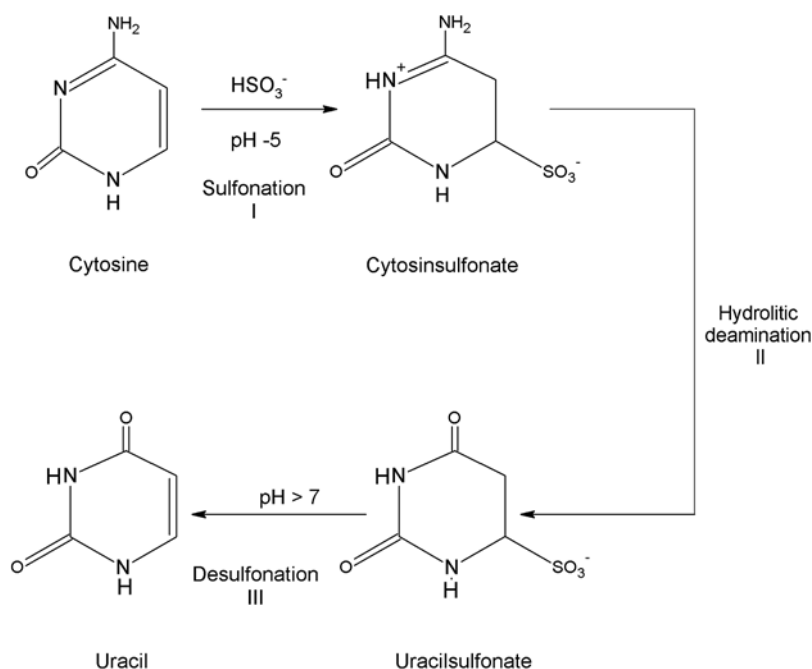


Fig. 1 Chemical reactions with the steps to convert unmethylated cytosines to uracils. Adapted from [19]

There have been a number of studies investigating the potential use of methylation differences for application to forensic serology. Lee and colleagues [16] selected five tDMRs to determine the DNA methylation status of pooled DNA samples of blood, saliva, semen, menstrual blood, and vaginal fluid and successfully identified two loci that can potentially discriminate semen from the other body fluids tested. However the experimental approach of using pooled and cloned DNA samples does not expose individual variation for each locus [17]. For that reason it is extremely important to study each locus in individual samples and determine if the differences observed in methylation are specific for the body fluid or just that individual.

The use of Pyrosequencing® for forensic determination of body fluids provides a robust method for analysis. The main advantage of the technique over previous methods is its ability to produce a quantitative determination of relative levels of methylation after bisulfite treatment (Fig. 1) at each targeted location. Individual Pyrograms® provide a set of methylation levels for each CpG site in the target region, and the precise level of methylation can be determined for multiple samples (Fig. 2). Madi and colleagues [18] screened a set of samples for different body fluids including saliva, sperm, blood, and skin epidermal cells. Using quantitative levels of methylation produced by Pyrosequencing, they were able to establish mean levels and the analytical precision for specific CpG sites, as well as for each tissue type tested (Table 1).

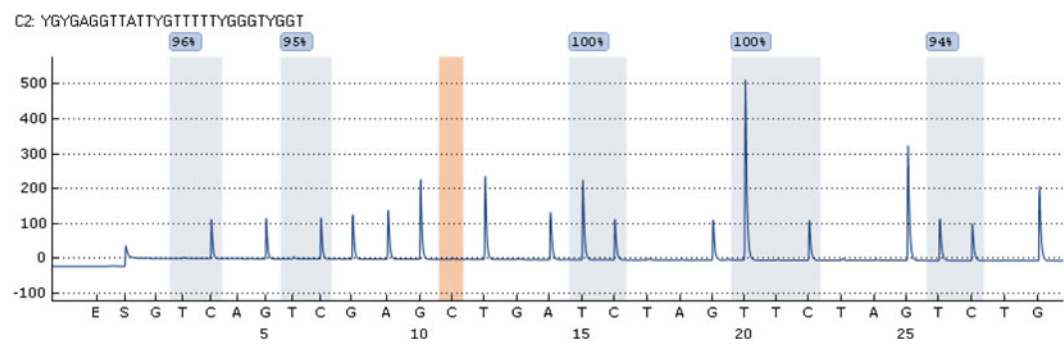


Fig. 2 Pyrogram showing the results for a blood sample sequenced for the locus *ZC3H12D*. The software calculates the percent methylation for each variable position (*blue shade*) according to the peak height ratio of the dispensed cytosine and thymine. The dispensation order for this sequence is shown above the pyrogram

Table 1
Mean percent of methylation and standard deviation for CpGs 1–5 for the *ZC3H12D* locus for all body fluids tested

Tissue type	Number of samples	Mean percent methylation ± SD				
		CpG1	CpG2	CpG3	CpG4	CpG5
Blood	10	94.9 ± 1.9	95.3 ± 1.6	100.0 ± 0.0	100.0 ± 0.0	85.5 ± 2.8
Saliva	11	86.6 ± 3.3	82.6 ± 4.8	100.0 ± 0.0	88.1 ± 4.1	82.6 ± 4.8
Skin cells	8	83.9 ± 30.3	94.3 ± 5.9	100.0 ± 0.0	83.3 ± 30.2	93.1 ± 5.5
Sperm	10	6.1 ± 1.2	5.5 ± 1.2	11.9 ± 1.9	8.3 ± 0.7	6.1 ± 1.2

Adapted from [18]

Overall the use of Pyrosequencing in body fluid determination is a promising technique for forensic laboratories. The added cost per sample for body fluid identification is offset by the high potential throughput of the methodology. Lastly, the addition of a body-fluid-specific determination to a DNA genotype presents important and sometimes vital information to the trier of fact in criminal casework.

2 Materials

2.1 DNA Extraction

1. Qiagen EZ1 DNA Investigator kit.
2. EZ1 Advanced Biorobot.
3. Buffer MTL.
4. Water bath or heating plate with controlled temperature.

2.2 Quantification

1. PicoGreen DNA reagent.
2. Deionized water molecular level or Milli-Q.
3. Real-time PCR thermal cycler.

2.3 Bisulfite Conversion

1. Epitect® Fast Bisulfite Conversion kit.
2. Ethanol 100 %.
3. Thermocycler.
4. Water bath or heated plate with controlled temperature.

2.4 Polymerase Chain Reaction

1. PyroMark® PCR Kit.
2. PyroMark CpG Assay BCAS4 (chr20:49,411,467-49,493,714).
Forward primer: AGTGGGTGAGGTTGTGAAATGT.
Reverse (biotinylated) primer: CCCATCCTACTAAAACATCTAATT.
Sequencing primer: AGTTT'TTTTGGTGAAGTTTAT.
3. PyroMark CpG Assay C20orf117 (chr20:35,412,178-35,492,087).
Forward, Reverse, and Sequencing primers are proprietary of Qiagen.
4. PyroMark CpG Assay FGF7 (chr15:49,715,375-49,779,523).
Forward (biotinylated) primer: GGGTTTATATGTATTATTTGGTGGT.
Reverse primer: CATTATATACTCCTCAAAACACACAC.
Sequencing primer: TATATACTCCTCAAAAC.
5. PyroMark CpG Assay ZC3H12D (chr6:149,768,766-149,806,148).
Forward, Reverse, and Sequencing primers are proprietary of Qiagen.

2.5 Agarose Gel

1. Agarose molecular biology grade.
2. TBE buffer 1×.
3. Ethidium bromide 0.5 µg/mL added on the agarose gel.
4. DNA ladder with good separation between the 150, 200, and 250 bp fragments such as Low Molecular Weight DNA Ladder from New England Biolabs.

2.6 Pyrosequencing: PyroMark® Q24 Advanced

- All materials are purchased from Qiagen.
1. PyroMark Q24 Advanced System.
 2. PyroMark Q24 Advanced Software.
 3. PyroMark Q24 Advanced CpG Reagents (4×24).

4. PyroMark Denaturation Solution (500 mL).
5. PyroMark Wash Buffer (concentrated, 200 mL).
6. PyroMark Q24 Cartridge [3].
7. PyroMark Q24 Plate (100).
8. PyroMark Q24 Plate Holder.
9. PyroMark Q24 Vacuum Prep Troughs.
10. PyroMark Q24 Vacuum Workstation (110 V).
11. PyroMark Vacuum Prep Filter Probe (100).

2.7 Streptavidin Beads

Streptavidin Sepharose High Performance Beads, 34 μm bead size in a highly cross-linked agarose matrix (GE Health Care Life Sciences, Piscataway, NJ).

3 Methods

3.1 DNA Extraction

1. With sterile scissors cut the swab into an appropriate tube.
2. Add 490 μL of a 1:2 dilution of buffer G2 and 10 μL of Proteinase K.
3. Vortex and incubate for more than 2 h at 56 °C. During incubation vortex the tube occasionally to improve cell lysis.
4. Using a spin basket remove the liquid from the swab.
5. All the 500 μL of lysate should be in the 2 mL sample tubes provided with the kit.
6. Add 400 μL of Buffer MTL to each sample tube with lysate. Vortex.
7. Load the sample tubes in the proper location of the EZ1 Advanced BioRobot (follow instruction manual).
8. Load all tubes and tips to the machine as well as the reagent cartridges (*see Note 1*).
9. Run the Large Volume protocol with 50 μL final volume elution with water.
10. Run the program. Remove the samples and freeze at -20 °C for large periods of time or at 4 °C if quantification is performed on the same day.

3.2 Quantification

1. Prepare DNA standards from 50 ng/ μL down to 1 ng/ μL concentration by serial dilutions of the stock.
2. Add 29 μL of molecular grade Millipore water to enough 0.5 mL tubes for all standards, no-template control, and samples.
3. Add 2 μL of each standard or sample to the appropriate tubes and 2 μL water to the control. Make a 1:10 and/or 1:2 dilution of each unknown sample to make sure that the calculated value is within the range for the standard curve.

4. In the dark add 31 μL of a 1:200 dilution of PicoGreen solution in DMSO to each tube.
5. Vortex, quick spin to remove drops from the inside of lids, and incubate for at least 5 min in the dark.
6. Transfer 20 μL aliquots of each tube to three different 0.2 mL (PCR) tubes.
7. Load tubes in the real-time PCR machine with the standard of highest concentration on position 1 (per manual instruction).
8. Run an isothermal fluorescent quantification protocol on a Rotor-gene system using settings for 20 μL of sample and hold for 2 min at 50 °C followed by ten cycles of 15 s at 50 °C with fluorescence reading after each cycle on FAM (470 nm source and 510 nm detector at gain 5).
9. Determine concentration of the sample based on the standard curve.

3.3 Bisulfite Conversion

1. Samples and all reagents need to be at room temperature.
2. According to each sample's concentration pipette enough volume to have 500 ng of DNA to a 0.2 mL (PCR) tube.
3. Add molecular grade water up to 40 μL of volume.
4. Add 85 μL of bisulfite solution.
5. Add 15 μL of DNA protect buffer.
6. Vortex and quick spin on the centrifuge to remove drops from inside lid.
7. Place tubes in a thermal cycler and run the program 5 min at 95 °C, 20 min at 60 °C, 5 min at 95 °C, and 20 min at 60 °C and hold indefinitely at 20 °C (*see Note 2*).
8. Briefly centrifuge the tubes and transfer the solution to 1.5 mL tubes.
9. Mix carrier RNA solution with buffer BL at the ratio of 3.5 μL carrier RNA to 350 μL Buffer BL. Multiply each volume for the number of samples.
10. Add 310 μL of the mixture buffer BL with carrier RNA to each tube. Vortex and quick spin in the centrifuge.
11. Add 250 μL of ethanol (96–100 %) to each sample. Vortex for 15 s and quick spin in the centrifuge.
12. Transfer the content of each tube to labeled MinElute® DNA spin columns.
13. Centrifuge the columns for 1 min at 12,000 $\times g$ or 13,000 rpm. Discard the liquid on the bottom of tubes (flow through) and place the columns back on the collection tubes.
14. Invert Buffer BW several times before use and add 500 μL of Buffer BW to each spin column. Centrifuge for 1 min at

- 12,000 $\times g$ or 13,000 rpm, discard the flow through, and place the columns back on the collection tubes.
15. Add 500 μL of Buffer BD to each tube; close the tubes and incubate for 15 min at room temperature; make sure that Buffer BD is not open for longer than required.
 16. Centrifuge the columns for 1 min at 12,000 $\times g$ or 13,000 rpm. Discard the flow through and place the spin columns back in the collection tubes
 17. Add 500 μL of Buffer BW to each spin column and centrifuge for 1 min at 12,000 $\times g$ or 13,000 rpm. Remove flow through and place spin columns back in the collection tubes.
 18. Repeat **step 17**.
 19. Add 250 μL of ethanol (96–100 %) to each spin column and centrifuge for 1 min at 12,000 $\times g$ or 13,000 rpm.
 20. Place the spin columns in new 2 mL collection tubes. Centrifuge for 1 min at 12,000 $\times g$ or 13,000 rpm to remove any remaining liquid.
 21. Incubate the spin columns for 5 min at 60 °C with the lids open to evaporate any residual ethanol.
 22. Place the spin columns into clean labeled 1.5 mL tubes (not provided with the kit).
 23. Add 10 μL of Buffer EB into the center of the membrane on the spin column. Incubate for 1 min at room temperature and centrifuge for 1 min at 12,000 $\times g$ or 13,000 rpm.
 24. Repeat **step 23** (*see Note 3*).
 25. Store the converted DNA at –20 °C.

3.4 Polymerase Chain Reaction

1. Thaw the PyroMark PCR Master Mix, the Coral Load Concentrate, molecular grade water aliquot, aliquot of the PyroMark assay primers, and bisulfite-modified DNA to be amplified.
2. To determine the number of reactions, add the number of DNA samples with 2 for no-template controls (NTC, no DNA is added, just water) plus one (to account for pipetting loss).
3. Prepare a master mix by adding 12.5 μL of PyroMark PCR Master Mix, 2.5 μL of Coral Load Concentrate, 2.5 μL of PyroMark assay primers, and 5.5 μL of water per reaction. To determine the volumes to add, multiply each volume given per the number of reactions determined in **step 2**.
4. Aliquot 23 μL of the mix prepared in **step 3** to the appropriate number of 0.2 mL (PCR) tubes.
5. Add 2 μL of molecular grade water to the NTC tubes and 2 μL of DNA to the remaining tubes.

6. Centrifuge the tubes and load them in a thermal cycler. Program the thermal cycler for an initial step of 15 min at 95 °C followed by 45 cycles of 30 s at 94 °C, 30 s at 55 °C, and 30 s at 72 °C, with a 10-min final extension at 72 °C and a last hold at 4 °C for infinite time (*see Note 4*). For custom primers the annealing temperature is usually set to $T_m - 5$ °C.
7. When the PCR is complete, remove the tubes and store at -20 °C.

3.5 Agarose Gel

1. Prepare a 2 % agarose gel in 1× Tris-borate-EDTA buffer.
2. Cool the gel and add ethidium bromide before pouring the gel in the tray.
3. Let the gel solidify for at least 20 min.
4. Take the comb out and place the tray with the gel on the electrophoresis apparatus.
5. Load the ladder with the loading buffer and 2–5 µL each PCR product to the appropriate wells on the gel. The PCR product does not need loading buffer since the Coral Load Concentrate has that function.
6. Turn the electrophoresis power ON at a proper current (10 V/cm) for the size of the gel.
7. Take a picture of the gel under an ultraviolet light and check for the bands around 200 bp size and no bands for the NTC.

3.6 Pyrosequencing Q24 Advanced System

1. Turn the Pyrosequencer ON at least 30 min prior to use.
2. Thaw the PCR products, an aliquot of biotinylated PCR primers (this aliquot is only used for Pyrosequencing and should never be used to amplify to avoid contamination from PCR products to pre-amplification DNA), and the sequencing primer. Equilibrate at room temperature the diluted washing buffer, denaturation buffer, annealing solution, binding solution, streptavidin-coated beads, and PyroMark Gold reagents. Place the metal tray for the Q24 plate on a heating plate set at 80 °C.
3. Prepare the bead mixture by adding 40 µL of binding buffer, 29 µL of molecular grade water, and 1 µL of beads per reaction. Multiply each volume per the number of samples plus one for pipetting loss. Add 70 µL of this mixture to 22 wells of a 24-well PCR plate. Leave the last two wells empty for negative controls and do not put bead mixture in these wells (*see Note 5*).
4. Add 10 µL of PCR product in the corresponding well of the 24-well PCR plate. Close the wells with a strip cap and mix at 1,400 rpm for at least 15 min (*see Note 6*). Remove from mixing only when the vacuum station is ready to use.

5. Add 20 μL of 1 $\times$ sequencing primer in annealing buffer to 22 wells on the PyroMark Q24 sequencing plate. The last two wells are reserved for the primer controls. Add the sequencing primer in one well, and add a mixture of biotinylated PCR primer (2 μL) and sequencing primer (20 μL) to the other well.
6. Place the PyroMark Q-24 plate on the vacuum station and fill all the trays of the station with the corresponding solutions.
7. Prepare the template for the plate on the Pyrosequencing software. Check the volumes to add to the cartridge for each nucleotide, enzyme mix, and substrate mix. Add the volumes to the cartridge and place it on the Q24 Pyrosequencer.
8. Turn the vacuum switch in the suction probe to “ON” and place it on the prime tray with distilled water aspirating approximately 70 mL.
9. Remove the 24-well PCR plate from agitation, take the lids off (*see Note 7*), and place in the vacuum station. Place the suction probe into the PCR plate wells with the vacuum position “ON” carefully capturing all the solution from the wells for about 15 s.
10. Wash the beads/samples with 70 % ethanol for 5 s.
11. Continue to the tray with denaturing solution, and flush for 5 s.
12. Aspirate for 10 s on the washing buffer tray and tilt the suction probe to beyond vertical 90° for a few seconds.
13. Place the vacuum suction probe on top of the PyroMark® Q24 plate without touching the liquid. Turn off vacuum and then lower the handle on the PyroMark® Q24 sequencing plate. Shake the suction probes gently from side to side for 30 s to 1 min to release the beads on the sequencing plate containing sequencing primers.
14. Rinse the vacuum suction probe on the rinse tray containing distilled water by agitating for 10 s.
15. Aspirate distilled water on the prime tray with aspirating 70 mL of water.
16. Move the suction probe beyond 90° vertical for a few seconds and disconnect the vacuum.
17. Place the sequencing plate on the metal tray and heat at 80 °C for 5 min.
18. Within a 30-s interval place the sequencing plate on the Q24 Pyrosequencer, close the lid, and start the run.
19. When the run is complete discard the Q24 plate and the unused contents of the cartridge by inverting it and rinse the cartridge with deionized water. Fill up each well of the cartridge with water and by applying pressure on the top with

your fingers force the water to pass the small needles on the bottom. Repeat this at least three times for each well. Let it dry at room temperature.

20. Remove and rinse all trays with deionized water from the vacuum station and dry at room temperature.

3.7 Data Analysis

1. The PyroMark Q24 software will perform the methylation percent analysis at each variable CpG sites and the data is displayed as a pyrogram.
2. The user should look up the target sequence on a genome browser online and select appropriate bisulfite controls to check the bisulfite conversion of “C” that are not part of the CpG site.
3. The bisulfite controls are analyzed by dispensing a cytosine in a position where the unconverted sequence has a cytosine but the bisulfite-converted sequence has a thymine (when dealing with the forward strand).
4. If any warning is displayed based on incomplete bisulfite conversion the whole assay should be discarded.
5. If any yellow or red warning is displayed due to a height variation on a homopolymer, deselect that peak as a reference peak.
6. If red warnings are displayed due to high peak to peak height variation the affected variable positions are not considered and the assay should be repeated.
7. The percent methylation value for each variable position is copied to an Excel spreadsheet with values for all tissues tested for each marker.
8. For unknown samples the tissue of origin is determined by comparing the methylation values for the unknown with the published values [18] for each marker.

4 Notes

1. The reagent cartridges should be loaded after all tubes and tips and right before the beginning of protocol to avoid having the magnetic beads settle on the bottom of the cartridge. After this the rack containing the lysed cells and the empty tubes and tips must be placed in the correct spot, the door of the EZ1 closed, and the program started. Note that the EZ1 software in the machine comes with its own instructions on the loading of the tubes, tips, and cartridges. One needs to manually accept all steps so that the run can start.
2. The Qiagen protocol suggests the possibility of increasing the incubation time at 60 °C to 20 min to guarantee a complete bisulfite conversion. We have not measured an increase in

DNA fragmentation or any other problems; therefore, we adopted this time as standard protocol in order to avoid incomplete conversion that would give false results in the Pyrosequencing analysis.

3. To improve yield we add twice the 10 μ L of elution buffer, which is the minimum amount recommended by Qiagen. We also perform a 1-min incubation step after each addition since we believe that it will improve yield. These adaptations to the Fast Epitect Bisulfite modification kit were adopted from a kit being previously used in our laboratory—the Epitect Bisulfite modification kit.
4. The default annealing temperature on the PyroMark PCR kit is 56 °C. We performed an optimization of the PCR protocol by evaluating which temperatures within the range of 54–58 °C would give more PCR product. Although the results for the temperatures of 55 and 56 °C were close to each other, we decided to use 55 °C as annealing temperature. We recommend that a similar procedure be performed for all new primers that will be adopted as markers for body fluid determination to ensure a good yield of PCR product. Also we do not use the Q solution that comes with the kit because it diminished the PCR efficiency for the samples that were tested in our laboratory.
5. A 96-well non-skirted PCR plate can be obtained commercially and cut into a 24-well plate format (8 $\times$ 3) to fit in the vacuum station.
6. The shaker must be able to shake at 1,400 rpm and hold 24-well PCR plates. In addition, a 2 mm orbital diameter is preferable to other shakers.
7. The lids of the PCR plate must be removed with caution to avoid the contamination of drops from one reaction to the next well of the plate. A compromise between tapping the plate in the bench top to make drops on the lid to go down and avoiding the delay of starting the vacuum procedure must be found to avoid the beads from depositing on the bottom of the plate.

References

1. Butler J (2010) Sample collection, storage and characterization. In: Butler J (ed) *Fundamentals of forensic DNA typing*, 1st edn. Elsevier, New York
2. Zubakov D, Hanekamp E, Kokshoorn M et al (2008) Stable RNA markers for identification of blood and saliva stains revealed from whole genome expression analysis of time-wise degraded samples. *Int J Legal Med* 122: 135–142
3. Haas C, Hanson E, Kratzer A et al (2011) Selection of highly specific and sensitive mRNA biomarkers for the identification of blood. *Forensic Sci Int Genet* 5:449–458
4. Hanson EK, Ballantyne J (2013) Highly specific mRNA biomarkers for the identification of vaginal secretions in sexual assault investigations. *Sci Justice* 53:14–22
5. Juusola J, Ballantyne J (2003) Messenger RNA profiling: a prototype method to supplant

- conventional methods for body fluid identification. *Forensic Sci Int* 135:85–96
6. Lindenbergh A, de Pagter M, Ramdayal G et al (2012) A multiplex (m)RNA-profiling system for the forensic identification of body fluids and contact traces. *Forensic Sci Int Genet* 6:565–577
 7. An JH, Shin K, Yang WI et al (2012) Body fluid identification in forensics. *BMP Rep* 45:545–553
 8. Juusola J, Ballantyne J (2005) Multiplex mRNA profiling for the identification of body fluids. *Forensic Sci Int* 152:1–12
 9. Juusola J, Ballantyne J (2007) mRNA profiling for body fluid identification by multiplex quantitative RT-PCR. *J Forensic Sci* 52: 1252–1262
 10. Rottach A, Leonhardt H, Spada F (2009) DNA methylation-mediated epigenetic control. *J Cell Biochem* 108:43–51
 11. Vidaki A, Daniel B, Court DS (2013) Forensic DNA methylation profiling-potential opportunities and challenges. *Forensic Sci Int Genet* 7:499–507
 12. Eckhardt F, Lewin J, Cortese R et al (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet* 38: 1378–1385
 13. Li C, Zhao S, Zhang N et al (2013) Differences of DNA methylation profiles between monozygotic twins' blood samples. *Mol Biol Rep* 40:5275–5528
 14. Song F, Mahmood S, Ghosh S et al (2009) Tissue specific differentially methylated regions (TDMR): changes in DNA methylation during development. *Genomics* 93:130–139
 15. Sliker RC, Bos SD, Goeman JJ et al (2013) Identification and systematic annotation of tissue-specific differentially methylated regions using the illumina 450 k array. *Epigenetics Chromatin* 6:26
 16. Lee HY, Park MJ, Choi A et al (2012) Potential forensic application of DNA methylation profiling to body fluid identification. *Int J Legal Med* 126:55–62
 17. An JH, Choi A, Shin K et al (2013) DNA methylation-specific multiplex assays for body fluid identification. *Int J Legal Med* 127:35–43
 18. Madi T, Balamurugan K, Bombardi R et al (2012) The determination of tissue-specific DNA methylation patterns in forensic biofluids using bisulfite modification and pyrosequencing. *Electrophoresis* 33:1736–1745
 19. Lilischkis R, Kneitz H, Kreipe H (2001) Methylation analysis of CpG islands. *Methods Mol Med* 57:271–283

Free ebooks ==> www.Ebook777.com

INDEX

A

ABCC6 116–119
 Absorbance 5, 6, 137
 Accuracy 42, 64, 90, 237, 272, 280, 281, 286, 364, 365
 AER. *See* Allelic expression ratio (AER)
 Alignment 53, 54, 117, 192, 193, 196, 340–343, 346, 354, 366
 Allele frequency 11, 43, 155, 278, 312
 Allele quantification (AQ) 22, 23, 75, 53–170, 243, 285
 Allele quantification (AQ) mode 23, 26, 92, 98, 165, 216, 250, 253, 254, 304, 312
 Allele-specific expression 243, 254
 Allele-specific methylation 167, 223, 226, 237, 271–287, 293, 304
 Allele-specific PCR (AS-PCR) 64, 84, 131, 137, 191, 292, 293, 300–302, 311
 Allelic expression 153–170, 242, 254
 Allelic expression ratio (AER) 166, 168
 Allelic imbalance 153–170
 Alu 201, 202, 305, 325
 Amniotic fluid 123, 124, 127, 128, 131
 Amplification bias 42, 43, 221, 222
 AMR. *See* Antimicrobial resistance (AMR)
 Analytical sensitivity 40, 45, 84, 272, 273, 280
 Aneuploidy 123–132, 285
 Angelman syndrome 242, 255, 261
 Antibody 50, 316, 318, 319, 323, 330, 331, 398
 Anti-EGFR targeted therapy 40
 Antimicrobial resistance (AMR) 364–366, 368, 373
 Apyrase 6, 8–10, 13, 18, 19, 125, 129, 338, 350
 Archival 104, 170
 ART. *See* Assisted reproduction technology (ART)
 AS-PCR. *See* Allele-specific PCR (AS-PCR)
 Assisted reproduction technology (ART) 260, 261
 ATP 5–9, 19, 30, 84, 91, 125, 127–129, 154, 163, 211, 212, 216–218, 292, 296, 317, 338, 350, 367, 382
 Autosomal DNA 155, 379, 381

B

Bacillus anthracis 364, 368
 Bacteria 11, 12, 127, 349, 357, 360
 Bacterial typing 11, 337–347, 349–361, 363–373

Baseline fluctuation 45
 Beckwith-Wiedemann syndrome 242, 260
 Betaine 10
 Bioenergetics 4, 5
 Bioluminescence 3, 125, 128, 153
 Biotinylation 159
 Bisulfite pyrosequencing 202–204, 206, 221–238, 315–332
 Bisulfite sequencing 210, 222, 273
 Blastocyst 259–269
 Blocking oligonucleotides 10
 Body fluid identification 400
 Bone marrow
 BRAF 37, 38, 63–80
 BRAF V600E 64–71, 76–79, 154

C

Calibration 42, 44, 52, 55–56, 98, 243
 Calibration curve 242, 253, 254
 cDNA 43, 53, 156, 158–160, 166, 168, 242, 244, 245, 247–249, 262, 264
 ChIP. *See* Chromatin immunoprecipitation (ChIP)
 Chromatin 190, 315, 316, 318, 319, 321–324, 330, 331
 Chromatin immunoprecipitation (ChIP) 315–332
 Circulating DNA 89, 90, 96
 CNV. *See* Copy number variation (CNV)
 COLD-PCR 83–100
 Colorectal cancer (CRC) 49–51, 54, 55, 64, 98, 181
 Copy number variation (CNV) 115–121
 COSMIC database 50, 71, 72, 89
 CpG dinucleotides 195, 271, 317
 CpG island 175, 179, 190, 217, 256, 271, 272, 275–280, 285, 317, 319, 398
 CpG mode 23, 26
 CpN sites 26
 CRC. *See* Colorectal cancer (CRC)
 Critical temperature 84, 85, 88, 94–96
 Cross-linking 402
 Cut-off values 285

D

dATP α S. *See* Deoxyadenosine alpha-thio triphosphate (dATP α S)
 7-deaza-2'-deoxyadenosine-5'-triphosphate 10

412 | PYROSEQUENCING: METHODS AND PROTOCOLS
Index

- Deoxyadenosine alpha-thio triphosphate (dATP α S).. 19, 91, 127, 163, 211, 215–218, 296, 367, 382
- Diagnostics 45, 49, 175, 225, 294, 320, 349
- Differentially methylated regions (DMRs) 244, 260, 261, 263, 266–268, 317, 398, 399
- DMRs. *See* Differentially methylated regions (DMRs)
- DNA quantification 39, 87–91, 99, 237, 244, 250–255, 278, 281, 294, 305, 320, 325, 329, 380, 401, 402
- Drug resistance 349, 350
- E**
- EGFR..... 38, 40, 49, 50
- E-*ice*-COLD-PCR 83–100
- eis*..... 351, 353
- EMA. *See* European Medicines Agency (EMA)
- Embryo..... 222, 226, 228, 232, 234, 236, 237, 259–261, 264–269, 271
- Endonucleases 209, 222
- Enrichment 83–86, 94–96, 98, 100, 292, 318, 324, 331
- Environment 131, 201, 208, 224, 236, 261, 398
- Epigenetics 12, 17, 22, 104, 176, 190, 209, 222, 237, 241, 242, 244, 255, 256, 259–261, 272, 304, 315, 398
- Epigenome 264, 271, 272
- Epigenomics..... 219, 288, 332
- Expression analysis 22
- F**
- False-negative results 64, 67, 107, 287
- False-positive results..... 107, 176, 272, 273, 275, 285
- FFPE tissue DNA..... 53, 60, 74, 87, 89, 179, 286, 287
- Food and Drug Administration (FDA)..... 64
- Forensic casework..... 379–397
- Forensics 11, 12, 379–408
- Formalin-fixed paraffin-embedded (FFPE) 44, 52, 53, 55, 56, 60, 64, 68, 69, 74, 76, 86–90, 96, 104, 106, 109, 167, 179, 286, 287, 312, 332
- Fungal typing 11
- G**
- GC content 27, 108, 179, 278, 355, 356, 360
- Genomic imprinting..... 154, 241, 242
- Global DNA methylation..... 201–206, 209–218
- gyrA* 351, 353
- H**
- Haplotyping 393
- Hematology..... 11, 104
- Heteroduplex..... 84, 85, 94
- Heteroplasmy 11, 380
- Heterozygote 129, 130, 165, 166
- High resolution meting (HRM)..... 83, 99
- Histone..... 190, 271, 315–319, 323, 330
- Histone modification..... 190, 271, 315–319, 330
- HLA. *See* Human leukocyte antigen (HLA)
- HLA typing..... 12
- Homopolymer .. 10, 23–26, 40, 107, 112, 124, 126, 129, 179, 185, 237, 257, 356, 381, 395, 407
- Hotspot mutation..... 50
- HRM. *See* High resolution meting (HRM)
- Human leukocyte antigen (HLA) 12, 133–150
- Hydroxymethylcytosine..... 211, 273
- Hypermethylation 175, 176, 190, 255, 256, 318
- Hypomethylation 176, 256, 261
- I**
- ice*-COLD-PCR..... 85
- ICR 244, 254, 255, 257
- Immunoprecipitation..... 315, 319, 323
- Imprint control region (icr)
- Imprinted gene 227, 228, 237, 242–244, 253, 255, 260, 261, 265, 291, 293
- Imprinting 154, 237, 241–258, 260, 272, 293, 317
- Indel 22, 26
- inhA*..... 351, 353
- Inhibitor 64, 133, 134, 272, 319, 322, 323, 394
- In vitro diagnostics 21, 23
- Isomer..... 9, 10
- Isoschizomer..... 209, 210, 217
- K**
- Karyotyping..... 123, 129
- katG*..... 351, 353
- Klenow enzyme 7
- KRAS..... 32–35, 38, 49–60, 86, 89, 95, 98, 99, 154
- L**
- Limiting dilution 221–238
- Limit of blank (LoB)..... 41
- Limit of detection (LoD) 29, 40–42, 84, 94–96, 100, 104, 176, 292, 317
- Limit of quantification (LoQ)..... 41–42
- LINE-1 201–206
- Linearity 42, 43, 90
- Locked nucleic acid (LNA)..... 85, 87–89, 92–94, 96, 292, 294, 298, 305, 307, 310, 320, 325
- LOH. *See* Loss of heterozygosity (LOH)
- LOI. *See* Loss of imprinting (LOI)
- Long range epigenetic silencing (LRES)..... 190
- Loss of heterozygosity (LOH) 153–170
- Loss of imprinting (LOI) 241–258

LUMA. *See* Luminometric Methylation Assay (LUMA)
 Luminometer..... 9, 217
 Luminometric methylation assay (LUMA)..... 209–218

M

Magnetic beads..... 6–8, 407
 Malignancy..... 63, 190, 275
 Melanoma 63–65
 Melting profile 282, 287
 Methylation level..... 40, 43, 45, 176, 184,
 190, 191, 196, 198, 201–206, 216, 235, 244, 246,
 250–257, 275, 285, 310, 311, 317, 318, 329,
 332, 399
 Methylation-specific PCR (MSP)..... 43, 176,
 210–218, 272, 273, 275, 277, 278, 280, 281, 283, 285,
 286, 293
 5-methylcytosine 209–218, 272, 273
 MethylLight..... 305, 325
 Methylome 187, 219
 MHC 134
 Microbial detection 364
 Mitochondrial DNA (mtDNA) 380, 381
 MPN. *See* Myeloproliferative neoplasms (MPNs)
 mPSQed..... 340–343, 345
 mRNA..... 53, 154, 155, 242, 261,
 264, 267, 398
 MSP. *See* Methylation-specific PCR (MSP)
 Multiplex pyrosequencing® 337–347, 392
 Multiprimer sequencing 10
 MultiPSQ..... 341, 345, 347
 Mutation detection..... 10, 37, 67, 84, 86, 88,
 97–99, 108
 Mutation hotspot 83, 86, 89–92, 95, 98, 99
 Mycobacteria 360
 Myeloproliferative neoplasms (MPNs) 103, 140

N

NRAS..... 49–60, 63, 64, 89

O

Oocyte..... 222, 226, 227, 234, 236, 237, 259, 269

P

Paralogous sequences..... 115–121
 PCDHB cluster..... 189–200
 Phi29 polymerase 44, 233
 Plasma 86, 88, 89, 96
 PMF. *See* Primary myelofibrosis (PMF)
 Pooled samples 11, 260
 Prader-Willi syndrome 255
 Precision 40, 42, 45, 64, 399
 Predictive therapy..... 49

Preimplantation..... 259–261, 264
 Prenatal diagnosis..... 123–132
 Primary myelofibrosis (PMF)..... 103
 Primer dimer 53, 91, 92, 94, 95, 99, 108, 126,
 224, 226, 232, 278, 279, 286, 296, 298, 326, 344, 357,
 360, 393, 394
 Processivity..... 12
 Pseudogene..... 54, 99, 115, 116, 190, 192
 Pyrogram®..... 19, 28–38, 67, 116, 140, 165,
 205, 304, 324, 329, 337, 355, 366, 381
 Pyromaker 30–38, 111
 PyroMark® Instruments..... 17–27
 Pyrophosphatase..... 6, 10

Q

Quality control 39–45, 86, 165, 185, 206,
 275, 277, 279, 305, 309, 317, 325

R

RAS-hypothesis 50
 Read length 10, 12, 13, 21, 23, 24, 110, 176,
 226, 293, 298, 327, 338, 351, 360, 365, 393
 Real-time PCR..... 45, 87, 89, 90, 93, 99, 222,
 275, 278–280, 282, 285, 294, 320, 365, 394, 401, 403
 Real-time sequencing 64, 84, 221, 292, 317
 Repetitive elements 201–206
 resistance 22, 40, 349–352, 364, 368, 373
 RNA 17, 104, 158, 214, 242, 244, 247,
 256, 264, 267, 292, 304, 305, 323, 329, 330, 398, 403
rpoB..... 351, 353
rrs 351–353

S

Sanger sequencing 3, 11, 29, 30, 32, 64–68,
 84, 104, 245, 247–248, 292, 317, 338, 347, 350, 351,
 364, 380, 392
 Secondary structures..... 8, 9, 86, 90, 126, 178,
 179, 224, 226, 232, 344, 355, 360, 395
 Sensitivity..... 5, 6, 21, 40, 45, 60, 64, 76, 84,
 96, 100, 202, 211, 217, 261, 272, 273, 280, 322, 350,
 351, 394
 Sepsis..... 12
 SEQ mode..... 24, 26
 Sequenase..... 7, 10
 Sequence analysis (SQA)..... 23, 369, 380
 Sex determination..... 380, 381, 390, 391
 Short tandem repeat (STR)..... 379–381, 383,
 385, 390, 392, 394, 395
 Silver-Russell syndrome 261
 Single cell analysis 222
 Single molecule analysis 223
 Single-nucleotide variations (SNV) 192

414 | PYROSEQUENCING: METHODS AND PROTOCOLS
Index

- Single stranded DNA binding protein (SSB)..... 9, 10, 395
 SNV. *See* Single-nucleotide variations (SNV)
 Solid-phase sequencing 6–8
 SQA. *See* Sequence analysis (SQA)
 SRSF2 104–106, 109, 112
 SSB. *See* Single stranded DNA binding protein (SSB)
 SssI methylase 44
 Standards 52, 55–56, 84, 87–91, 233, 273,
 280, 281, 285, 287, 324, 402
 STR. *See* short tandem repeat (STR)
 Sulfurylase 8, 19, 84, 125, 129, 211,
 212, 292, 317, 337, 338, 350
 Supratyping 133–150
- T**
- Termination-recognition base (TRB)..... 381, 390, 391
 Theranostics 49
 Tissue-specificity 397–408
 Transposable elements (TEs) 201, 202, 272
 TRB. *See* Termination-recognition base (TRB)
 Trisomy 21..... 124, 126, 130
 Tuberculosis..... 349–361
- Tumor cell percentage 30, 107
 Tumor heterogeneity 285
- U**
- U-BRAF^{V600} assay 63–79
- V**
- Validation 31, 39–45, 88, 93–94, 176, 224,
 280–281, 307, 309, 331, 332
 Vascular endothelial factor (VEGF)..... 50
 Virtual Pyrograms 32, 33, 36, 37, 111
 Virus..... 12
- W**
- Whole genome amplification (WGA)..... 99, 222,
 233, 273, 280, 286
 Workflow..... 19–21, 26, 86,
 213, 283, 284
- Y**
- Yersinia pestis*..... 368